

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 19, 2003, 14:59:25 ; Search time 28 Seconds  
(without alignments)  
861.776 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273  
Sequence: 1 LRLRRFEINVLVSGGAAGK.....RKADQIMRGYIKALEVLSE 251

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 73: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	934	73.4	259	2	F70420 conserved hypotnet
2	391.5	30.8	728	2	E83328 hypothetical prote
3	300.5	23.6	275	2	C83973 hypothetical prote
4	297.5	23.4	764	2	B82303 conserved hypotnet
5	276	21.7	293	2	AB3539 serine proteinase
6	274.5	21.6	300	2	A82017 probable lipoprote
7	273	21.4	312	2	AH2791 conserved hypotnet
8	273	21.4	314	2	AF3365 serine proteinase
9	273	21.4	315	2	H97570 hypothetical 34.4K
10	268.5	21.1	260	2	H69874 conserved hypotnet
11	268.5	21.1	297	2	A81002 conserved hypotnet
12	268.5	21.1	395	2	F82852 conserved hypotnet
13	258.5	20.3	314	2	G85703 hypothetical prote
14	258.5	20.3	314	2	H90845 hypothetical prote
15	257.5	20.2	301	2	AH0649 conserved hypotnet
16	257.5	20.2	314	2	B36871 probable membrane
17	253.5	19.9	1048	2	A70592 hypothetical prote
18	243	19.1	345	2	G83441 conserved hypotnet
19	242	19.0	1679	2	S49802 probable membrane
20	240	18.9	311	2	D71252 conserved hypotnet
21	235.5	18.5	610	2	B87518 conserved hypotnet
22	235	18.5	583	2	A70729 hypothetical prote
23	234	18.4	360	2	E70892 hypothetical prote
24	231.5	18.2	308	2	AE2661 conserved hypotnet
25	231.5	18.2	311	2	C97443 hypothetical prote
26	231	18.1	1065	2	A70797 hypothetical prote
27	223.5	17.6	291	2	F69959 hypothetical prote
28	208.5	16.4	1351	2	S44655 ZK170.4 protein -
29	199.5	15.7	296	2	E84000 hypothetical prote

30	189.5	14.9	860	2	T21745 hypothetical prote
31	162	12.7	283	2	H97198 probable phosphos
32	161	12.6	1316	2	T50444 hypothetical UPF00
33	153	12.0	254	2	G72343 conserved hypotnet
34	152.5	12.0	324	2	D70943 hypothetical prote
35	147.5	11.6	329	2	A87087 conserved hypotnet
36	147.5	11.6	598	2	E71657 hypothetical prote
37	142	11.2	356	2	E85663 hypothetical prote
38	142	11.2	356	2	G90803 hypothetical prote
39	142	11.2	687	2	T09051 hypothetical prote
40	139.5	11.0	288	2	T17654 probable aspartate
41	138.5	10.9	353	2	E87668 conserved hypotnet
42	138.5	10.9	749	2	S61643 probable membrane
43	134	10.5	801	2	T47774 hypothetical prote
44	134	10.5	825	2	T48431 hypothetical prote
45	133	10.4	356	2	F90978 hypothetical prote

## ALIGNMENTS

RESULT 1  
F70420  
conserved hypothetical protein aq\_1386 - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C/Accession: F70420  
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.  
V.  
Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: F70420  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-259 <AQR>  
A/Cross-references: GB:A800737; NID:92983782; PIDN:AA07357.1; PID:92983799; GB:AE0  
A/Experimental source: strain VF5  
C/Genetics:  
A/Gene: aq\_1386

Query Match 73.4%; Score 934; DB 2; Length 259;  
Best Local Similarity 71.5%; Pred. No. 9.9e-74;  
Matches 178; Conservative 31; Mismatches 40; Indels 0; Gaps 0;

QY	1	LRLRRFEINVLVSGGAAGKIAHIGVAKAINEIGIRVALSGVSAQIVSVFYASGYSPB	60
DB	5	LRLRRFEINVLVSGGAAGKIAHIGVAKAINEIGIRVALSGVSAQIVSVFYASGYSPB	64
QY	61	GNFSLKRVNMLKFKFPLKGLIGWEKAIKRLAEVLPRRIEKLIEPTTCATDLVYSG	120
DB	65	ENLKLKRVNMLKFKFPLKGLIGWEKAIKRLAEVLPRRIEKLIEPTTCATDLVYSG	124
QY	121	RALYVSEGSILPALGSCAIPGIFEPVEYKAYLLVVDGIVNNLPVEPFGESGIPYCVDV	180
DB	125	KALYVGRDILIPVLLGSCSIPGIFEPVEYKAYLLVVDGIVNNLPVEPFGESGIPYCVDV	184
QY	181	LPIERKQIKNIMLILASFLAVRSNKKKKCDLVYVPELAEFPPLDVRKADQIMER	240
DB	185	LPIERKQIKNIMLILASFLAVRSNKKKKCDLVYVPELAEFPPLDVRKADQIMER	244
QY	241	GYIKALEVL 249	
DB	245	GYESTLRIM 253	

RESULT 2  
E83328  
hypothetical protein PA339 [imported] - Pseudomonas aeruginosa (strain PA01).  
C/Species: Pseudomonas aeruginosa  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: E83328  
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

A: Map position: I



QY 61 -----GMSLLKRVNWLKLPKPKPLKGLIGWEKAIRLEEVLPYRIEKLPTIYCAT 115  
 Db 93 SLTRRRNFLLD-----ITFRG--SGLFGMKLDGRRLRHLOGIRIEDLPKFFVAVCT 143  
 QY 116 DLYSGRALYLSGSLIPALLGSCAIPGIFEPVEYKNYLLVDGGIVNNLPVE----- 166  
 Db 144 ELRTGHEIWLSTGLPVEAMRSYALPGVPEPVRMGDRVLDGALVNPVPSVCRAVBOAL 203  
 QY 167 -----PQESGPIPTVCVDVLPPIEPEKDIK-NILHILLRSFFLAVR 205  
 Db 204 VLAANLHYOYGRAAVIKAKRQESVPOA-----MGEXETRGITGVWMAFNIIQD 257  
 QY 206 SNSEKRR--BFCDLVIVPELEFTPLDVRKADQIMERGYIKALEVLSE 251  
 Db 258 RISRRMAGDPDVSIMPTVGQIGLADFHRAAEADAGYTETVKRIED 305  
 RESULT 9  
 H97570  
 hypothetical 34.4K protein in hmr-puru intergenic region [imported] - Agrobacterium tumefaciens  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
 C:Accession: H97570  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
 A:Reference number: A97359; PMID:11743194  
 A:Accession: H97570  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-315 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK87521.1; PID:gl5156852; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_3215  
 A:Map position: circular chromosome  
 Query Match 21.4%; Score 273; DB 2; Length 315;  
 Best Local Similarity 26.0%; Pred. No. 4e-16;  
 Matches 75; Conservative 47; Mismatches 84; Indels 82; Gaps 4;  
 QY 9 INVLSCGAAKGTAGHIGLVKAINELGIRVRLSGVSAAGIYVSVFVASCYSPGEMFSLK 68  
 Db 37 IALALGGAARGWAHIGVLRALDEAGVKIGMAGTSIGALVGGCYLAG-----K 85  
 QY 69 VNWMLKLPKPKPLK-----GLIGWEKAIRFLEEVLPYRIEKLPTIYCAT 115  
 Db 86 LDELEEFARSLLTWRIAGLLDLTIGGGGLFGGRLTKRMQHEGLRVENLEHFFIAVAT 145  
 QY 116 DLYSGRALYLSGSLIPALLGSCAIPGIFEPVEYKNYLLVDGGIVNNLPV----- 165  
 Db 146 ELRTGHEIWLSTGLPVEAMRSYALPGVPEPVRMGDRVLDGALVNPVPSVCRAVBOAL 205  
 QY 166 -----EPFOESGPIPTVCVDVLPPIEPEKDIK-NILH 194  
 Db 206 VVAVNMLNLDLGRSAVVVKAASPOGGTTPAETAPRGLPGVWVQAFNIIQDR----- 258  
 QY 195 ILLRSFFLAVRSNKKKFCDLAVIVPELEFTPLDVRKADQIMERGY 242  
 Db 259 -----ISKRLAGDPDPLMLHPRINDIGLSEFHRASEIDRGY 296  
 RESULT 10  
 H69874  
 conserved hypothetical protein ylbk - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: H69874  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertsch, C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emmermann, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gai, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, P.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardine, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Scandell, A.; Schleicher, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; S. Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, K.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, A.; Yoshikawa, H.P.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: H69874  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-260 <KUN>  
 A:Cross-references: GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13377.1; PID:el1850  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ylbk  
 Query Match 21.1%; Score 268.5; DB 2; Length 260;  
 Best Local Similarity 28.2%; Pred. No. 7.7e-16;  
 Matches 68; Conservative 54; Mismatches 106; Indels 13; Gaps 4;  
 QY 8 EINVLSCGAAKGTAGHIGLVKAINELGIRVRLSGVSAAGIYVSVFVASCYSPGEMFSLK 67  
 Db 5 KIGALSGGARGLHIGLVKAINELGIRVRLSGVSAAGIYVSVFVASCYSPGEMFSLK 60  
 QY 68 RVNWLKLPK-----FKPPLKGLIGWEKAIRFLEEVLPYRIEKLPTIYCATLYSGR 121  
 Db 61 KV--AKAFKRLVADYTPKLGDRVRLVHATPGKPIEELQIPGLGIVACDLQGTGE 118  
 QY 122 ALXLSGSLIPALLGSCAIPGIFEPVEYKNYLLVDGGIVNNLPVEPFOESGPIPTVCV-DV 180  
 Db 119 KIVFKGVSVDVARSISIGIFIPQRLDGLRLVDGAVDRIPVSVKMGADIIASDV 178  
 QY 181 LPPEKDIK-NILHLLRSFPLAVRSNKKKFCDLVIVPELEFTPLDVRKADQIMERGY 240  
 Db 179 SRVRKTETAVHIFDVIMQSMILQNLVRLHQTIAADIMIRPSLETYSSTSSSFANIEEMISA 238  
 QY 241 G 241  
 Db 239 G 239  
 RESULT 11  
 A81002  
 conserved hypothetical protein NMB2139 [imported] - Neisseria meningitidis (strain MC58)  
 C:Species: Neisseria meningitidis  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: A81002  
 R:Pettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiagiani, V.; Pizsa, M. Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755; PMID:10710307  
 A:Accession: A81002  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-297 <TET>  
 A:Cross-references: GB:AE002562; GB:AE002098; NID:g7227392; PIDN:AAF42447.1; PID:g722  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB2139  
 Query Match 21.1%; Score 268.5; DB 2; Length 297;  
 Best Local Similarity 31.4%; Pred. No. 9.1e-16;  
 Matches 70; Conservative 41; Mismatches 97; Indels 15; Gaps 3;  
 QY 9 INVLSCGAAKGTAGHIGLVKAINELGIRVRLSGVSAAGIYVSVFVASCYSPGEMFSLK 68



Db 43 VGLALGGAGSKGPAHAGIIVKENGIPKVVATGVSAGSIVGSLFASGMSDPRLERAEI 102

Qy 69 VNNLKLFRKPPKGLIGWEKARFLEEVLPYRIEKEIFPVYCATOLYSGRALVYSEG 128

Db 103 LKGTDLVDLTLSTSGIKKEKONINRKGROIQCPPIKFAVADPFGKAVANOG 162

Qy 129 SLIPALGSCAIPGIEPEYKNTLLVDGIVNNLPVEPQSGIPYCVDLPIPEKD 188

Db 163 MGOQVAFASAIPIVFPQPIGRHYVDGSLQPPVPSAARQGANFYIAVDISARPKN 222

Qy 189 IKNIILHLRSPF-----LAVRSNKKRKEF--CDLIVPEL 223

Db 223 IS-----QGFFSYLDQTLNMTSVSALONEGADVIXQV 258

RESULT 12

F82852

conserved hypothetical protein XF0066 [imported] - Xylella fastidiosa (strain 9a5c)

C/Species: Xylella fastidiosa

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C/Accession: F82852

R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences

Nature 406, 151-157, 2000

A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A/Reference number: A82515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: F82852

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-355 <STM>

A/Cross-references: GB:AE003860; GB:AE003849; NID:99104830; PIDN:AF82879.1; GSPDB:GN001

A/Experimental source: strain 9a5c

R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, R.; Docena, C.; El-Dorry, H.; Facincani, A.F.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klegger, J.E.; Kurame, E.B.; Laigh chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E. A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Mitracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sattelli, R.V.; Savasch A.; Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A/Reference number: A59328

A/Contents: annotation

C/Genetics:

A/Gene: XF0066

Query Match 21.1%; Score 268.5; DB 2; Length 395;

Best Local Similarity 29.4%; Pred. No. 1.3e-15;

Matches 74; Conservative 48; Mismatches 113; Indels 17; Gaps 4;

Qy 9 INDVLSGAAKGIHAGVILKAINELGIRVALSGVSAIYVFAAGSYSEGGFSLIKR 68

Db 96 IGTALGGAAKGPANHGVLMLEANGIAPSVVAGSAGSVGALYASGMD---SFKQTK 152

Qy 69 -----VNNLKLFRKPPKGLIGWEKARFLEEVLPYRIEKEIFPVYCATDIYSG 120

Db 153 AVAMDESNIRDLRFPS-----GGLVQOQKLOMYVNELVGKPIREKLPFAIAITRLDGD 207

Qy 121 RALVLSGSLIPALGSCAIPGIEPEYKNTLLVDGIVNNLPVEPQSGIPYCVDLPIPEKD 179

Db 208 QRTMFYKAVGVAVASCSICGVPEPVITGAYHYDGSITSPVVDANROIGADPIVAVD 267

Qy 180 VLPIPEKQIKNIILHLRSPFLAVRSNKKRKEFCDLIVPELEETPLDVRRAQIME 239

Db 268 ISKATKRNPGGIVGVNQSISIMQOHGEALKRAIVIHPIKVDIGSTDFDORNAITL 327

Qy 240 RGYIKALEVISE 251

Db 328 EGERATVIVIRE 339

RESULT 13

G85703

hypothetical protein ychK [imported] - Escherichia coli (strain O157:H7, substrain E

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C/Accession: G85703

R/Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma iller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85460; MUID:21074935; PMID:111206531

A/Accession: G85703

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-314 <STG>

A/Cross-references: GB:AE005174; NID:912514952; PIDN:AAG56091.1; GSPDB:GN00145; UWGP A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Gene: ychK

Query Match 20.3%; Score 258.5; DB 2; Length 314;

Best Local Similarity 25.3%; Pred. No. 7.3e-15;

Matches 78; Conservative 49; Mismatches 106; Indels 75; Gaps 8;

Qy 3 LRKPEINLVSGAAGIAHIGVILKAINELGIRVALSGVSAIYVFAAGSYSEGGFSLIKR 62

Db 14 MKKI-KIGALGSGAARQSHIGVINALKVGLEIDIVAGCSIGSLVGAAYAC-----DR 67

Qy 63 FSLKRVNMLKFRKPPK-----GLIGWEKARFLEEVLPYRIEKEIFPVYCATDIYSG 113

Db 68 ISALF--DWVTSFSGYVDLRLMDLSWQGGILNGERFVQYREIMETENCSRFPAAV 125

Qy 114 ATDLYSGAALYSGSLIPALGSCAIPGIEPEYKNTLLVDGIVNNLPVEPQSGIPYCV 173

Db 126 ANLSTGRLMFTGDLHLAIRASCSTIPGLMAPVANHGWLVGAVVNPPIPSLTRALGA 185

Qy 174 PTVCDVPIPEKQIKNIILHL-----LRSF----- 200

Db 166 DIVIN-----VDLQDHMLMOQDLSTFVNSSENSDISLPIWALRKEIGSTTTR 237

Qy 201 -----FLAVRSNKKRKEFCDLIVPELEETPLDVRRAQIMEGYI 243

Db 238 AVTAPLATEIMTTSIOVLNENLKRNMAQDPDILIOVPCPOISTIDFRAHAAIAAGOI 297

Qy 244 KALEVISE 251

Db 298 AVEKQMD 305

RESULT 14

H90845

hypothetical protein ECs1736 [imported] - Escherichia coli (strain O157:H7, substrain

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C/Accession: H90845

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, ( gaawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and A/Reference number: A99629; MUID:21156233; PMID:11258796

A/Accession: H90845

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-314 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA835159.1; PID:G13361201; GSPDB:GN00154 A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Gene: ECs1736

Query Match 20.3%; Score 258.5; DB 2; Length 314;

Best Local Similarity 25.3%; Pred. No. 7.3e-15;

Search completed: June 19, 2003, 15:04:47  
Job time : 29 secs

```

Matches 78; Conservative 49; Mismatches 106; Indels 75; Gaps 8;
QY 3 LRKEEINVLGSGAAGIAHIGVLKAINELGIRVALSGVSAGAIVSFVAGSYSEGM 62
Db 14 MRKI-KIGLALSGAARGNSHIGVINALKKGIEIDIVAGCSIGSLVGAAYAC-----DR 67
QY 63 FSLKRVNMLKFKFKPLK-----GLIGWEKAIKFLVLEVLVYRRIKLEIPTYIC 113
Db 68 LSALE--DWTFSYMDVLRMLDLSWQGGLLRGVFNQYREIMPETEINCSRRFAAV 125
QY 114 ATDLYSGHALYSEGLIPALLGSCAIPGIFEPVEYKYNLLVDGGIVNNLFPVEFQESGI 173
Db 126 ATNLSTGRELWFTGDLHLAIRASCISFGLMAPVAHNGYWLVDGAVVNPVPSLTRAIGA 185
QY 174 PTVCVDVLPTEPEKDKNHLIL--LRSF----- 200
Db 186 DIVIA-----VDLQDAHLMOQDLSLNFVNSENGDSLPWHARKLERIGSIYTRR 237
QY 201 -----FLAVRSNSEKKEFC-----DLVIVPELEEFPLDVRKADQIMERGYI 243
Db 238 AVTAPTATEIMTTSIQVLENRLKRNRMAGDPDILIQVPCQISTLDFRAHAAIAAGQL 297
QY 244 KALEVLSE 251
Db 298 AVEKKMDE 305

RESULT 15
AH0649
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AH0649
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
  T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
  S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AP0502; PMID:11677608
A:Accession: AH0649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <PA3>
A:Cross-references: GH:AL513382; PIDN:CAD08379.1; PID:gl6502423; GSPDB:GN00176
C:Genetics:
A:Gene: STY1296

Query Match 20.2%; Score 257.5; DB 2; Length 301;
Best Local Similarity 26.3%; Pred. No. 8.4e-15;
Matches 79; Conservative 49; Mismatches 113; Indels 59; Gaps 7;
QY 3 LRKEEINVLGSGAAGIAHIGVLKAINELGIRVALSGVSAGAIVSFVAGSYSEGM 62
Db 1 MRKM-KIGLALSGAARGNSHIGVINALKKGIEIDIVAGCSIGSLVGAAYAC-----NK 54
QY 63 FSLKRVNMLKFKFKPLK-----GLIGWEKAIKFLVLEVLVYRRIKLEIPTYIC 113
Db 55 LSALEQ--WVCFSYMDVLRMLDLSWQGGLLRGVFNHRYDIMEVDFDHCRRFGAV 112
QY 114 ATDLYSGHALYSEGLIPALLGSCAIPGIFEPVEYKYNLLVDGGIVNNLFPVEFQESGI 173
Db 113 ATNLSTGRELWFTGDLHLAIRASCISFGLMAPVAHNGYWLVDGAVVNPVPSLTRAIGA 172
QY 174 PTV-----CVDVLPTEPEKDKNHLILKRSF----- 201
Db 173 DIVIAVDLQDAHLMOQDLSLNFVNSENGDSLPWHARKLERIGSIYTRR 232
QY 202 -----LAVRSNSEKKEFC-----DLVIVPELEEFPLDVRKADQIMERGYIKALEVLSE 251
Db 233 EIMTTSICVLENRLKRNRMAGDPDILIQVPCQISTLDFRAHAAIAAGQLAVEKKMDE 292

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 14:21:10 ; Search time 44 Seconds  
(without alignments)  
760.134 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273  
Sequence: 1 LRRLKFEINLVLSGNAAG.....RKADQIEMERYIKALEVLSR 251

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.GeneSeq.101002.\*  
1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1273	100.0	251	18	AAW23071 Aquifex pyrophilus
2	913.5	71.8	249	18	AAW23074 Aquifex esterase V
3	269.5	21.2	300	20	AAV38779 Neisseria meningit
4	268.5	21.1	300	20	AAV38778 Neisseria meningit
5	259.5	20.4	300	20	AAV38780 Neisseria gonorrhoe
6	223.5	17.6	1327	21	AAV70474 Human cyclic nucle
7	217.5	17.1	1389	22	ABBS5863 Drosophila melanog
8	188.5	14.8	259	18	AAW23079 Whale mat sample 1
9	186	14.6	1402	22	ABG22116 Novel human diapo
10	168	13.2	283	23	ABP28552 Streptococcus poly

11	168	13.2	283	23	ABP29892
12	146	11.5	356	22	ABBS2902
13	145	11.4	149	20	AAV38777
14	142	11.2	269	23	ABP30534
15	136	10.7	282	22	AAW25456
16	120	9.4	501	22	AAU61400
17	118	9.3	507	22	ABW71378
18	109.5	8.6	504	22	AAV53641
19	109.5	8.6	504	22	AAW70159
20	107.5	8.4	543	20	AAW86149
21	107	8.4	480	22	ABW61534
22	96.5	7.6	341	22	ABG22110
23	96.5	7.6	376	21	AAW38929
24	96.5	7.6	381	21	AAW38928
25	96.5	7.6	484	21	AAW38927
26	94.5	7.4	492	23	ABW91378
27	89.5	7.0	361	21	AAW21352
28	89.5	7.0	367	21	AAW21351
29	89.5	7.0	376	21	AAW21350
30	89.5	7.0	388	23	ABW5483
31	88.5	7.0	361	21	AAW38048
32	88.5	7.0	367	21	AAW38047
33	88.5	7.0	373	21	AAW38046
34	88.5	7.0	388	18	AAW26740
35	88	6.9	349	23	ABW54471
36	88	6.9	838	22	ABG22113
37	87	6.8	404	21	AAW30647
38	87	6.8	414	21	AAW30646
39	86.5	6.8	288	23	ABW54054
40	86	6.8	1161	21	AAV52037
41	86	6.8	1161	21	AAV51666
42	86	6.8	1161	23	ABW84776
43	85.5	6.7	546	22	AAW39046
44	85.5	6.7	558	22	AAW38969
45	84.5	6.6	384	21	AAW61569

## ALIGNMENTS

RESULT 1					
AAW23071	ID	AAW23071	standard; Protein; 251 AA.		
XX					
AC	AAW23071;				
XX					
DT	17-FEB-1998	(first entry)			
XX					
DE	Aquifex pyrophilus esterase 28LC.				
XX					
KW	Esterase; thermostable enzyme; ester; chiral compound; cheese;				
KW	pulp; paper; lignin removal; sugar; lignocellulose;				
KW	disease resistance; feedstuff.				
XX					
OS	Aquifex pyrophilus strain KO1 5a.				
XX					
FH	Key	Location/Qualifiers			
FT	Misc-difference 248				
FT		/note= "encoded by TGA"			
XX					
PN	MO9730160-AI.				
PD	21-AUG-1997.				
XX					
PF	11-FEB-1997;	97WO-USO2039.			
XX					
PR	16-FEB-1996;	96US-0602359.			
XX					
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.				
XX					
PI	Callen W, Kosmicka A, Link S, Maffia AM, Murphy D;				
XX	Reid J, Robertson DE, Swanson RV, Warren PV;				
XX					

Streptococcus poly  
Escherichia coli p  
Neisseria meningit  
Streptococcus poly  
Human protein sequ  
Propionibacterium  
Drosophila melanog  
A bone marrow secr  
DNA encoding human  
Fat cell different  
Drosophila melanog  
Novel human diapo  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Herbicideally activ  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human sperm skelem  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Straphylococcus car  
Lactococcus lactis  
Novel human diapo  
Arabidopsis thalia  
Arabidopsis thalia  
Lactococcus lactis  
A. aeolicus AAE05  
A. aeolicus AAE05  
DNA polymerase III  
Human polypeptide  
Human polypeptide  
Arabidopsis thalia

DR WPI: 1997-425035/39.  
 DR N-PSDB; AAT79324.  
 XX Nucleic acid encoding heat stable esterase from thermophilic  
 PT bacteria - which is active in organic solvents, useful in cheese or  
 PT paper manufacture, and to study plant resistance to disease  
 XX  
 XX Claim 19; Page 56-57; 113pp; English.  
 XX  
 CC This protein comprises the thermostable esterase 28LC of Aquifex  
 CC pyrophilus, a Gram-negative, strictly chemolithoautotrophic knall  
 CC gas marine bacterium which grows optimally at 85 deg C and pH  
 CC 6.8. Newly identified polynucleotides (AAT79321-30) encoding claimed  
 CC esterases (AAW23069-77, AAW23088) were recovered from genomic gene  
 CC libraries. They can be used for recombinant production of the  
 CC enzymes in host cells, and as probes to identify related sequences.  
 CC The esterases are stable at high temperature and in organic  
 CC solvents, making them superior for use in production of pure chiral  
 CC compounds used in pharmaceutical, agricultural and other chemical  
 CC industries. A method is claimed for transferring an amino group  
 CC from an amino acid to an alpha-keto acid using a claimed esterase.  
 CC The enzymes may also be useful as ripening starters in cheese making,  
 CC in lignin removal in paper and pulp manufacture, in carbohydrate  
 CC derivative synthesis, in fermentable sugar production from  
 CC lignocellulosic waste, in the study of plant wall structure, plant  
 CC resistance to disease and organic matter decomposition and to select  
 CC plants bred for production of highly degradable animal feeds.  
 XX  
 SQ Sequence 251 AA;

Query Match 100.0%; Score 1273; DB 18; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-133;  
 Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LRLKFEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSAIAVSVFYASGYSP 60  
 DB 1 LRLKFEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSAIAVSVFYASGYSP 60  
 QY 61 GMSILKRVNWLKLPKFPKPLKGLIGWEKAIKLEEVLPYRIEKLPTTICATDLYSG 120  
 DB 61 GMSILKRVNWLKLPKFPKPLKGLIGWEKAIKLEEVLPYRIEKLPTTICATDLYSG 120  
 QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 180  
 DB 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 180  
 QY 181 LPIEPEKDKNHLHILIRSFPLAVRSNKKRKFCDLVVPELEBFTPLDVRKADQIMER 240  
 DB 181 LPIEPEKDKNHLHILIRSFPLAVRSNKKRKFCDLVVPELEBFTPLDVRKADQIMER 240  
 QY 241 GYIKALVLS 251  
 DB 241 GYIKALVLS 251

RESULT 2  
 AAW23074  
 ID AAW23074 standard; Protein: 249 AA.  
 XX  
 XX AAW23074;  
 AC  
 DT 17-FEB-1998 (first entry)  
 XX  
 DE Aquifex esterase VF5-34LC.  
 XX  
 KW Esterase; thermostable enzyme; ester; chiral compound; cheese;  
 KW pulp; paper; lignin removal; sugar; lignocellulose;  
 KW disease resistance; feedstuff.  
 XX  
 OS Aquifex sp. strain VF5.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 99

PT XX /note= "encoded by TAA"  
 FN WO9730160-A1.  
 XX 21-AUG-1997.  
 PD  
 XX 11-FEB-1997; 97WO-US02039.  
 PF  
 XX 16-FEB-1996; 96US-0602359.  
 PR  
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 FA  
 XX Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;  
 PI Reid J, Robertson DE, Swanson RV, Warren PV;  
 XX WPI: 1997-425035/39.  
 DR N-PSDB; AAT79327.  
 XX  
 CC Nucleic acid encoding heat stable esterase from thermophilic  
 CC bacteria - which is active in organic solvents, useful in cheese or  
 CC paper manufacture, and to study plant resistance to disease  
 CC  
 CC Claim 19; Page 59-60; 113pp; English.  
 CC  
 CC This protein comprises the thermostable esterase VF5-23LC of  
 CC Aquifex VF5, a marine strictly chemolithoautotrophic knall  
 CC gas bacterium that grows optimally at 85 deg C and pH 6.8.  
 CC Newly identified polynucleotides (AAT79321-30) encoding claimed  
 CC esterases (AAW23069-77, AAW23088) were recovered from genomic libraries.  
 CC They can be used for recombinant production of the enzymes in host  
 CC cells, and as probes to identify related sequences. The esterases  
 CC are stable at high temperature and in organic solvents, making them  
 CC superior for use in production of pure chiral compounds used in  
 CC pharmaceutical, agricultural and other chemical industries. A  
 CC method is claimed for transferring an amino group from an amino  
 CC acid to an alpha-keto acid using a claimed esterase. The enzymes  
 CC may also be useful as ripening starters in cheese making, in lignin  
 CC removal in paper and pulp manufacture, in carbohydrate derivative  
 CC synthesis, in fermentable sugar production from lignocellulosic  
 CC waste, in the study of plant wall structure, plant resistance to  
 CC disease and organic matter decomposition and to select plants bred  
 CC for production of highly degradable animal feeds.  
 XX  
 SQ Sequence 249 AA;  
 Query Match 71.8%; Score 913.5; DB 18; Length 249;  
 Best Local Similarity 72.0%; Pred. No. 8.5e-93;  
 Matches 177; Conservative 31; Mismatches 37; Indels 1; Gaps 1;  
 QY 1 LRLKFEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSAIAVSVFYASGYSP 60  
 DB 5 LKLRFEVNLVLSGGAAGIAHIGVLKALEELGKVKRLSGVSAIAVSVFYASGYTPD 64  
 QY 61 GMSILKRVNWLKLPKFPKPLKGLIGWEKAIKLEEVLPYRIEKLPTTICATDLYSG 120  
 DB 65 EMLKLLKEVNLKLPKFPKPLKGLIGWEKAIKLEEVLPYRIEKLPTTICATDLYSG 124  
 QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 180  
 DB 125 KALYFCRGDLIPVLLGSCSIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 184  
 QY 181 LPIEPEKDKNHLHILIRSFPLAVRSNKKRKFCDLVVPELEBFTPLDVRKADQIMER 240  
 DB 185 LPITQERKKNHLHILIRSFPLAVRSNKKRKFCDLVVPELEBFTPLDVRKADQIMER 243  
 QY 241 GYIKAL 246  
 DB 244 GDMRAL 249

RESULT 3  
 AAY38779  
 ID AAY38779 standard; Protein: 300 AA.

XX	AAV38779;
AC	(first entry)
DT	08-OCT-1999
DE	Neisseria meningitidis strain A antigen encoded by ORF137.
KM	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
OS	Neisseria meningitidis.
PN	MO9924578-A2.
PD	20-MAY-1999.
PR	09-OCT-1998; 98WC-IB01665.
PR	01-SEP-1998; 98GB-0019016.
PR	06-NOV-1997; 97GB-0023516.
PR	14-NOV-1997; 97GB-0024190.
PR	18-NOV-1997; 97GB-0024386.
PR	27-NOV-1997; 97GB-0025158.
PR	10-DEC-1997; 97GB-0026147.
PR	14-JAN-1998; 98GB-0000759.
PA	(CHIR-) CHIRON SPA.
PI	Grandi G, Masignani V, Piazza M, Rapuoli R, Scarlato V;
DR	WPI, 1999-327407/27.
DR	N-PSDB; AA212214.
PT	Proteins from Neisseria meningitidis and N. gonorrhoeae useful for diagnosis, treatment and prevention of infection
PS	Claim 4; Page 324; 524pp; English.
CC	Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis and N. gonorrhoeae antigenic proteins. They are encoded by open reading frames (ORFs) AA21972-Z13358. The antigenic proteins, their fragments, their nucleic acids and antibodies are used for diagnosis, prevention (as vaccines) or treatment of Neisseria infections, such as meningitis, septicaemia and gonorrhea. Both organisms are closely related. Fragments of the nucleic acids are useful as hybridisation probes and antisense reagents.
SQ	Sequence 300 AA;
Query Match	21.2%; Score 269.5; DB 20; Length 300;
Best Local Similarity	31.4%; Pred. No. 3.9e-21;
Matches	70; Conservative 40; Mismatches 98; Indels 15; Gaps 3
Dy	9 INLVSGSAKGIAGIHLGKALNEIGIVRALSGVSAGATVSVFYASGVSPBGMSFLNR 68
Dd	46 VGALGGGAOKGFRAHVGIIKLKENGIDIPKVYVTGSAGSIYGLSIFASGMSPDLBLEARI 105
Dy	69 VMNLKLEKFPPEPKLIGMEKARILEBEVLPPREIELELPPTYICADLYSGALYLSRG 128
Dd	106 LGRKTLDVDLTLSISGFIKGEKIQNTYINRKVGGRRIQQPFKEAVALDFETGSAFAFNQS 165
Dy	129 SLPLPALLGSCAIGIFEVEYKNYLLVDGIVNNLPVPFOESGIPTVCADVLPJEPKO 188
Dd	166 NAQOAVAKASAIINVFOPVITGRHYTVVGGLSQPVPASAARRXXXXXVAIVDISAPSKN 225
Dy	189 IKNIILHLRSFF-----LAVRNSSEKRKEF--CDLVYPEL 223
Dd	226 IS-----QGFRSYLDQTLNMWSVASLQNELGQADVIVIKQV 261
RESULT 4	
AAV38778	
ID	AAV38778 standard; Protein; 300 AA.

XX AAAY8778; AC  
XX 08-OCT-1999 (first entry) DT  
XX Neisseria meningitidis antigen encoded by ORF137. DE  
XX Neisseria meningitidis, Neisseria gonorrhoeae; vaccine; KM  
XX treatment; Neisseria infection; meningitis; septicemia; gonorrhea. OS  
XX Neisseria meningitidis.  
XX WC0924578-AZ. FN  
XX PD 20-MAY-1999. PD  
XX 09-OCT-1998; 98MO-IB01655. XP  
XX 01-SEP-1998; 98GB-0019016. PR  
XX 06-NOV-1997; 97GB-0023516. PR  
XX 14-NOV-1997; 97GB-0024190. PR  
XX 18-NOV-1997; 97GB-0024386. PR  
XX 27-NOV-1997; 97GB-0025158. PR  
XX 10-DEC-1997; 97GB-0026147. PR  
XX 14-JAN-1998; 98GB-0000759. PR  
XX (CHIR-) CHIRON SPA. PA  
XX Grandt G, Masiagnani V, Piazza M, Rappuoli R, Scarlato V; P1  
XX WPI; 1999-327407/27. XX  
XX N-PSDB; AAZ12213. DR  
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for PT  
XX diagnosis, treatment and prevention of infection PT  
XX Claim 4; Page 323; 524pp; English. PS  
CC Amino acid sequences AAAY38499-Y38944 represent Neisseria meningitidis CC  
CC and N. gonorrhoeae antigenic proteins. They are encoded by open CC  
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins, CC  
CC their fragments, their nucleic acids and antibodies are used for CC  
CC diagnosis, prevention (as vaccines) or treatment of Neisseria CC  
CC infections, such as meningitis, septicaemia and gonorrhea. Both CC  
CC organisms are closely related. Fragments of the nucleic acids CC  
CC are useful as hybridisation probes and antisense reagents. CC

SQ Sequence 300 AA:

Query Match 21.1%; Score 268.5; DB 20; Length 300;  
Best Local Similarity 31.4%; Pred. No. 5.1e-21;  
Matches 70; Conservative 41; Mismatches 97; Indels 15; Gaps 3

QY 9 INLVSGGAAKGIHAHIGVLKAINELRVRALSGSVSAGAIIVSFVASGYSPGGMSFLKR 68  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
DQ VEGALGSGASKSQFAHVGIITKVLENGRIPVKVVTSSTGSIVGSLPASGMSPRBLEAETI 105  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
QY VMNMLKIFKKPPLKGLIGMERKAIRPLEEVLPYRIEELPTTICMNDLYSGRALYLSRG 128  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db LGKTLDVDLTLTSGFIKEKLONTNRKVGROIQPFKPAVAATDETISGAFAENG 165  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
QY SLIPALLSCAIPGIEPEVEKYKNYLVDGIVNNLPVEPOESGITPYVCVDVLPLEPKD 188  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db MACQAARAPAAAPNVFOPIVTIGRHYYDGISQEPVSAAAROGANFVIADVISAPEGKN 225  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
QY IKNIILHLRSFF-----LAVRSNSERKKEF--CDLIYVEL 223  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 226 IS-----QGFFSYLDQTIANWWSVALQNELGADVVIRQV 261  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :

RESULT 5  
AAAY38780  
AAAY38780 standard; Protein; 300 AA

XX AC AAY38780;  
 XX DT 08-OCT-1999 (first entry)  
 XX DE Neisseria gonorrhoeae antigenic protein encoded by ORF137.  
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 XX OS treatment; Neisseria infection; meningitis; septicaemia; gonorrhoea.  
 XX PN Neisseria gonorrhoeae.  
 XX PD W09924578-A2.  
 XX PF 20-MAY-1999.  
 XX PR 09-OCT-1998; 98WO-IB01665.  
 XX PR 01-SEP-1998; 98GB-0019016.  
 XX PR 06-NOV-1997; 97GB-0023516.  
 XX PR 14-NOV-1997; 97GB-0024190.  
 XX PR 18-NOV-1997; 97GB-0024386.  
 XX PR 27-NOV-1997; 97GB-0025158.  
 XX PR 10-DEC-1997; 97GB-0026147.  
 XX PR 14-JAN-1998; 98GB-0000759.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX PR WPI; 1999-327407/27.  
 XX DR N-PSDB; AA212215.  
 XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 XX PT diagnosis, treatment and prevention of infection  
 XX PS Claim 4; Page 325; 524pp; English.  
 XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
 CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,  
 CC their fragments, their nucleic acids and antibodies are used for  
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
 CC infections, such as meningitis, septicaemia and gonorrhoea. Both  
 CC organisms are closely related. Fragments of the nucleic acids  
 CC are useful as hybridisation probes and antisense reagents.  
 XX SQ Sequence 300 AA;  
 Query Match 20.4%; Score 259.5; DB 20; Length 300;  
 Best Local Similarity 31.4%; Pred. No. 5.1e-20;  
 Matches 70; Conservative 40; Mismatches 98; Indels 15; Gaps 3;  
 QY 9 INVLGGCAAGIAHIGVLKAINELGIRVRLSGVSAIIVSYAGSYGEGMFLSKR 68  
 Db 46 VALALGGGASKGFAHIGVIVKENGIPKVVVTGTSAGSIVGSLLAGSPDRLEAEI 105  
 QY 69 VNMVLEKFKPKLGLCKWKAIRFLERVLRYRIEKLPTTCATDLYSGRALVLSG 128  
 Db 106 LKTDVLVLTSTSGFIKGEKQYINRKVGRIQQFPKFAVATDFTGKAVAFNQ 165  
 QY 129 SLIPALLGSCAIPGIFEPVYKYNLLVDGIVNNLVPVFPQESGIPVVDVLPFPEKD 188  
 Db 166 NAGQAVRASAAIPNVQFVIIGRHKYVDGGLSQPVVSAARRQCANFVIAVDISARPSKN 225  
 QY 189 IKNILHILLRSPF-----LAVRSNSEKKEF--CDLVIVPEL 223  
 Db 226 VG-----QGFFSYLDQTLNMSVSVLQNELGQADVIVKPVQ 261  
 RESULT 6  
 AAY70474  
 ID AAY70474 standard; Protein; 1327 AA.

XX AC AAY70474;  
 XX DT 04-JUL-2000 (first entry)  
 XX DE Human cyclic nucleotide-associated protein-2 (CNAP-2).  
 XX KW Cyclic nucleotide-associated protein-2; CNAP-2; human; cytostatic;  
 KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;  
 KW immunomodulatory; anti-asthmatic; anti-anemic; anti-diabetic; diagnosis;  
 KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;  
 KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;  
 KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;  
 KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;  
 KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;  
 KW neurological; vision; reproductive; smooth muscle.  
 XX OS Homo sapiens.  
 XX Key Location/Qualifiers  
 XX Peptide 1..34  
 XX Protein 35..1327  
 XX /label= Mature CNAP-2  
 XX /note= "Shares 24% identity to Aquifex pyrophilus  
 XX esterase 28LC"  
 XX Modified-site 68  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1225  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 73  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 125  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 220  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 326  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 357  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 386  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 400  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 432  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 455  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 560  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 600  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 780  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 784  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 997  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1113  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1121  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1171  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1251  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1274  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1285  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1299  
 XX /note= "Potential phosphorylation site"  
 XX Modified-site 1301  
 XX /note= "Potential phosphorylation site"



PS Disclosure; SEQ ID NO 3681; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB161176-AB161175), expressed DNA sequences (AB161176-AB161175) and the encoded proteins (AB161176-AB161175).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1389 AA;

Query Match 17.1%; Score 217.5; DB 22; Length 1389;  
Best Local Similarity 26.1%; Pred. No. 2.2e-14;  
Matches 73; Conservative 44; Mismatches 108; Indels 55; Gaps 9;

QY 9 INLVSGGAAGIAHIGVLKAINELGIRVRLSGVSAAGVSVFYASGYSPGCMFSL-- 66  
DB 914 IGLVLGGGARGAAGHIGKAIQAGIPDMVGVSIGALMGWCS-----ERNITVTQ 969

QY 67 -----KRVNW-LKLFKPKPLKGLIGWEKAIKRLFEVLPRYRIEKLPTTYCATDLY 118  
DB 970 KARENSKMTKWFLLQDLDTVPTSMFGSGREFNKTIDTGDVSDIEDLWIPYFTLTDT 1029

QY 119 SGRALYSEGLIIPALLGSCAIPGIFEPV--EYKNYLLVDGGIVNNLPVFPFQESGIP-T 175  
DB 1030 ASCHRIHTGSLWRYRVSMSLSGMYPLCPDKGHLLDGGVYNNLPADVMMNLGAHI 1089

QY 176 VCDV-----LP1EPEKDIKNILHLIRSPFLAV 204  
DB 1090 IADVGSQDQDITLNYGDDLSGWLLYKKNWPFPSVKVPDLPIQSL-----AVSCV 1144

QY 205 RNSSE-KRKEFCDLVIVPELEFPLDVRKADQIMERGI 243  
DB 1145 RLEEVKNSDYCE-VIRPIDKYKTLAPGSPFDIRDVGIV 1183

RESULT 8  
AAW23079  
ID AAW23079 standard; Protein; 259 AA.  
XX  
AC AAW23079;  
XX  
DT 17-FEB-1998 (first entry)  
XX  
DE Whale mat sample 11.801 esterase es9.  
XX  
KW Esterase; thermostable enzyme; ester; chiral compound; cheese;  
KW pulp; paper; lignin removal; sugar; lignocellulose;  
KW disease resistance; feedstuff.  
XX  
OS Whale mat sample 11.801.  
XX  
PN WO9730160-A1.  
XX  
PD 21-AUG-1997.  
XX  
PF 11-FEB-1997; 97WO-US02039.  
XX  
PR 16-FEB-1996; 96US-0602359.  
XX  
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.  
XX  
PI Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;  
PI Reid J, Robertson DE, Swanson RV, Warren PV;  
XX  
DR WPI; 1997-425035/39.  
DR N-PSDB; AAT79332.  
XX

PT Nucleic acid encoding heat stable esterase from thermophilic  
PT bacteria - which is active in organic solvents, useful in cheese or  
XX paper manufacture, and to study plant resistance to disease  
PS Disclosure; Page 64-65; 113pp; English.

XX This protein comprises the whale mat sample 11.801 esterase es9.  
CC Newly identified polynucleotides (AAT79321-40) encoding esterases  
CC (AAW23069-88), some of which are claimed, can be used for  
CC recombinant production of the enzymes in host cells. The esterases  
CC are stable at high temperature and in organic solvents, making them  
CC superior for use in production of pure chiral compounds used in  
CC pharmaceutical, agricultural and other chemical industries. A  
CC method is claimed for transferring an amino group from an amino acid  
CC to an alpha-keto acid using a claimed esterase. The enzymes may  
CC also be useful as ripening starters in cheese making, in lignin  
CC removal in paper and pulp manufacture, in carbohydrate derivative  
CC synthesis, in fermentable sugar production from lignocellulosic  
CC waste, in the study of plant wall structure, plant resistance to  
CC disease and organic matter decomposition and to select plants bred  
CC for production of highly degradable animal feeds.

XX Sequence 259 AA;

Query Match 14.8%; Score 188.5; DB 18; Length 259;  
Best Local Similarity 27.3%; Pred. No. 3.2e-12;  
Matches 69; Conservative 43; Mismatches 118; Indels 23; Gaps 9;

QY 7 BEINLVSGGAAGIAHIGVLKAINELGIRVRLSGVSAAGVSVFYASGYSPGCMFSL 66  
DB 8 ENSLVLSGGGALGIAHIGVLHDLKQNIQVPEIIVGTSWGGIGASMAIGMKERILEBI 67

QY 67 KR-----VNMKLFKPKPLKGLIGWEKAIKRLFEVLPRYRIEKLPTTYCATDLYSG-- 120  
DB 68 KFSNVNENWIK-TSPSG--NSVVDNEKIAKIFDTLFKDKMTDTYIPLKLIATNLNGHK 124

QY 121 RALYLSGSLI-PALLGSCAIPGIFEPVRYKNVLLVDGGIVNNLPVFPFQESGIPVCDV 179  
DB 125 KVFTASDDVLKDAIILSTWAIQGVFEHIIDGETYDGFCLNGLVNEATFNDV--LAVD 182

QY 180 VL-----PIEPEKDIK--NILHLIRSPFLAVRNSSE---KRKEFCDLVIVPELEFPT 228  
DB 183 VMGENSPEKAMPDNFFKTSNVLEMFESKMLFIYNTQTHIKNANKNIYLIPTVKEYKT 242

QY 229 LDVRKADQIMERG 241  
DB 243 FQPHKHKIRALG 255

RESULT 9  
ABG22116  
ID ABG22116 standard; Protein; 1402 AA.  
XX  
AC ABG22116;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #22107.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX



PA (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX MPI: 2001-639362/73.  
 DR N-PSDB: AHS86303.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 20; SEQ ID No 52475; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 1402 AA;  
 SQ  
 Query Match 14.6%; Score 186; DB 22; Length 1402;  
 Best Local Similarity 25.5%; Pred. No. 7,1e-11;  
 Matches 74; Conservative 42; Mismatches 116; Indels 58; Gaps 10;  
 QY 3 LKKEBINLVSGG--AKGIAHIGVLAINEIGIRVALSGVAGAIIVSVFVAGYS-- 58  
 DB 965 LTRGNTLVVLEGRGARGGSHIGVLAIRBAGVVDLVGIGTIGFVIGVLEERSAS 1024  
 QY 59 -----PEGMFSILKRVNMLKLFKKKPLKGLIGMEKAIKRLAEVLP--YRRIEQL 108  
 DB 1025 RIKQARERAKSMSTVLEPV-----LDLTDAITSMTGSAANRSIHVFFKSRFEDFL 1079  
 QY 109 PTY-ICATDLYSGRALYLSGSL--IPALLGSCA--IPGIFPEVEYKNYLLVDGIVNN 162  
 DB 1080 PYFKKXPTDITASARGHDSGLMRYVRAWMOTLSGYLPPLCDPKD--GHILMDGGYINN 1137  
 QY 163 LVEPRQESGIPVGVVDVLPLEPEKIDNIIHLIRSFVAVRNS----- 208  
 DB 1138 LPADIRSMGAKTVAIDVGSODETDLSTYGDLSFGMWLLMKRLNPMGLTRXRPDMAET 1197  
 QY 209 -----EKRRKPCDVLIVPELEEFPTLDRKADIMERGY 242  
 DB 1198 QPFGPNVSCVQLAEVVKSSSYCE-YLPPIDCFMTQDFGFDQIDVGVY 1246  
 RESULT 10  
 ABP28552  
 ID ABP28552 standard; Protein; 283 AA.  
 XX  
 AC ABP28552;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DB Streptococcus polypeptide SEQ ID NO 6280.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 PA (GENO-) INST GENOMIC RES.  
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tetteilin H;  
 DR MPI: 2002-352536/38.  
 DR N-PSDB: AHN69183.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 PS Claim 1; Page 3791; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), AHN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 CC  
 XX Sequence 283 AA;  
 SQ  
 Query Match 13.2%; Score 168; DB 23; Length 283;  
 Best Local Similarity 23.7%; Pred. No. 6,8e-10;  
 Matches 66; Conservative 52; Mismatches 118; Indels 42; Gaps 8;  
 QY 9 INLVSGGAAGIAHIGVLAINEIGIRVALSGVAGAIIVSVFVAGYSPEGMFSILK 68  
 DB 6 VGLVLEGGGKGLTAGVLAFAIDAGIKIDGIVSVSGALFGVAVFSRORERL-----R 60  
 QY 69 VMVLKLFKKKPLKGLIGMEKAIKRLAEVLPYRRI-----EKLEIFTYICATD 116  
 DB 61 YN--KKYISHPKVMSLSRSMFRFTGNFVNKDFTYIVVPMKLVDPDEAKKSSIDFYVATE 118  
 QY 117 LYSGRALYLSGSL--IPALLGSCAIPGIFPEVEYKNYLLVDGIVNNLPVEPQESGI 173  
 DB 119 MTSGAPREYFKIDSVEBQMETLRASALPVVSKWMDQCKYLDGSLDIPVFAKGLG 178  
 QY 174 -PTVCVDVLPLEPEKIDN--IIHLIRSFVAVRNSERKRCFD-LIVPELEEFPTL 229  
 DB 179 DKLIVVMTRPLNPKQKSSGRLTYIKRYPNFKTASNNKYQYNNLSLEKVMLEKTKGL 238  
 QY 230 -----DVRKADIMERGYIKALAEVISE 251

Db 239 FAIRPSKSLVIGRLKPNPKLDSIYQLGMDAKSVMP 276

RESULT 11

ABP29892

ID ABP29892 standard; Protein; 283 AA.

XX

AC ABP29892;

XX

DT 02-JUL-2002 (first entry)

XX

DE Streptococcus polypeptide SEQ ID NO 8960.

XX

XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX

OS Streptococcus agalactiae.

XX

PN WO200234771-A2.

XX

PD 02-MAY-2002.

XX

XX 29-OCT-2001; 2001WO-GB04789.

XX

XX 27-OCT-2000; 2000GB-0026333.

PR

PR 24-NOV-2000; 2000GB-0028727.

PR

PR 07-MAR-2001; 2001GB-0005640.

XX

XX (CHIR-) CHIRON SPA.

PA

PA (GENO-) INST GENOMIC RES.

XX

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

PI

PI Tettelin H;

XX

XX WPI: 2002-352536/30.

DR

DR N-PSDB; ABN70523.

XX

XX New Streptococcus protein for the treatment or prevention of infection

PT

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX

XX Claim 1; Page 4017; 4525pp; English.

XX

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC

CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX

XX

SQ Sequence 283 AA;

Query Match 13.2%; Score 168; DB 23; Length 283;

Best Local Similarity 23.7%; Pred. No. 6.8e-10;

Matches 66; Conservative 52; Mismatches 118; Indels 42; Gaps 8;

9 INVLGGAAGKGIHGVKALNELGIRVRLSGVAGAVSVFVAGVSPGGMFSLKR 68

6 VGLVLEGGMEGLYTAGVLDALFDLADGKINDIGVSVAGALFGVNFVSRERAL-----R 60

69 VNWLLKFLFKPKPLKGLIGWEKAIPLFLEVLPRRI-----EKLRIPTYCATD 116

Db

QY

61 YN--KKYLSHPKYMSLRSWFTGCFNFKDFTYVEVPMKLDVDFDDEAFKSSIDFYVATE 118

117 LYSGRALYLSGSL----IPALLGSCAIFGIFEPVEYKRYLLVDGGIYVNNLPVRFQBSGI 173

119 MTSKPEYFKIDSVPFQMEILRASALPVWSKMDWQKKYLDGGLSDSIPEVDFARGLGF 178

174 -PTVCVDVLPFIEPEKDIKN--ILHILLRSFFLAVRSNSEKKEKFCFCD-LVIVPELEETPL 229

179 DKLIVVMTRPLNYQKKPSSGRLYKTKRYPNFVKTNRYQYNNLSLEKVMSEKTKGDL 238

230 -----DVRKADQIMERGVIKALEVLS 251

239 FAIRPSKSLVIGRLKPNPKLDSIYQLGMDAKSVMP 276

RESULT 12

ABB52902

ID ABB52902 standard; Protein; 356 AA.

XX

AC ABB52902;

XX

DT 11-FEB-2002 (first entry)

XX

XX Escherichia coli polypeptide SEQ ID NO 1215.

XX

XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;

XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;

XX systemic infection; non-diarrhoeal infection; septicemia;

XX pyelonephritis; antibiotic resistance.

XX

OS Escherichia coli.

XX

XX WO200166572-A2.

PN

XX

PD 13-SEP-2001.

XX

XX 12-MAR-2001; 2001WO-EP03445.

PF

XX

XX 10-MAR-2000; 2000FR-0003145.

PR

PR 02-FEB-2001; 2001FR-0001449.

XX

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX

XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

PI

PI WPI: 2001-550253/61.

DR

XX

XX A library of DNA fragments of Escherichia coli strains for the

PT

PT phylogenetic determination of a given strain comprises polynucleotides of

PT nature B2/D+ A- -

XX

XX Example 6; Fig 6; 646pp; English.

XX

XX The invention relates to a library of DNA fragments of Escherichia coli

CC

CC strains comprising polynucleotides (ABA89577-ABA89729 and ABA89533)

CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature

CC B2/D+A-. The polynucleotides have potential antiinflammatory,

CC antibacterial and immunosuppressive activity as part of pharmaceutical

CC compositions used to treat, palliate or prevent extra-intestinal E. coli

CC infections. The polypeptides are useful for determining the phylogenetic

CC group of a given E. coli strain. These polypeptides can detect and treat

CC an undesired development of E. coli, particularly an extra-intestinal

CC infection that include systemic and non-diarrhoeal infections such as

CC septicemia, pyelonephritis and meningitis this is particularly

CC advantageous as bacterial resistance is increasing with the more

CC frequent use of broad spectrum antibiotics.

XX

XX Sequence 356 AA;

Query Match 11.5%; Score 146; DB 22; Length 356;

Best Local Similarity 24.4%; Pred. No. 2.6e-07;

Matches 87; Conservative 40; Mismatches 92; Indels 138; Gaps 18;

QY 11 LVISGGAAGIAHIGVLKALNEIGIRVRLSGVSGAIVSYVYASGYSP----- 60  
 DB 9 LVISGGAGVAYQAGVVKALAECCGTQISMVSGASIGAFNGAIIAA--SPDLSEAAVRLTA 66  
 QY 61 -----GMFSLIKRVNMLKFEK-----PPLKG-----LIGW 87  
 DB 67 LMDHLKMNQVLSVRLVYFESLK-----KLFQAMNLCQIFGRAGALLTLRLHISLNGF 121  
 QY 88 EKAIKRLBEVLPRRIKLE--IPYICATDYSGRALYLS-----ECSL----- 130  
 DB 122 DNLM-----AQPLLSDEPLTALDHLTDALADGILPIYVSLYPTGGMQDIIICIRAE 176  
 QY 131 -----IP-----ALGSCAIPGIFEPVYKYNLLVDSG-----IYNLL 163  
 DB 177 GAGTTKNAVFOHTQSLPRGQKRALASALPLLRPREVGCTMGDGMGRMNGNT 236  
 QY 164 PVEPQSGIPYCV-----DVLPIEPEKDIK-----NIIHLIRSP 200  
 DB 237 PVTEPLVDAGCMVIVTHLSGSLMDRQAFPDYTLIEIRPKRLKXAGDNGSGGL--SF 294  
 QY 201 FLAVRSSEKKEKPCDLVIVLELEPTPLDVK--ADOIMERGYI--KALEVLSF 251  
 DB 295 TLA-HTDAMRQGGEDTMLME-----HIKPLAAQVLSRSETVLQSLSETER 343

## RESULT 13

AY38777  
 ID AAY38777 standard; Protein; 149 AA.

XX AAY38777;

DT 08-OCT-1999 (first entry)

DE Neisseria meningitidis antigen encoded by a partial ORF137.

KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 treatment; Neisseria infection; meningitis; septicemia; gonorrhea.

OS Neisseria meningitidis.

PN W09924578-A2.

PD 20-MAY-1999.

PF 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;

XX WPI; 1999-327407/27.

XX N-PSDB; AAZ1212.

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for

diagnosis, treatment and prevention of infection

Claim 4; Page 322; 524pp; English.

CC organisms are closely related. Fragments of the nucleic acids  
 CC are useful as hybridisation probes and antisense reagents.

XX Sequence 149 AA;

Query Match 11.4%; Score 145; DB 20; Length 149;  
 Best Local Similarity 50.0%; Pred. No. 9.5e-08;  
 Matches 26; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 9 INLVSGGAAGIAHIGVLKALNEIGIRVRLSGVSGAIVSYVYASGYSP 60  
 DB 46 VGLALGGASKGPAHVGIKVLKENGIVPVKVTGTSAGSIYGNLFSGMSPD 97

## RESULT 14

ABP30534  
 ID ABP30534 standard; Protein; 269 AA.

XX ABP30534;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 10244.

KM Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;

KM group A streptococcus; Streptococcus pyogenes; antibacterial;

KM antinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN W0200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Frazer C;

XX Tectelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABN71165.

XX New Streptococcus protein for the treatment or prevention of infection

or disease caused by Streptococcus bacteria, such as meningitis, and

for detecting a compound that binds to the protein.

Claim 1; Page 4155; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

the specification. The proteins have antibacterial and anti-inflammatory

activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and

antibodies that bind (1) are used in the manufacture of medicaments for

the treatment or prevention of infection or disease caused by

SQ	Sequence	269 AA;	
XX	Query Match	11.2%; Score 142; DB 23; Length 269;	
PS	Best Local Similarity	22.8%; Fred. No. 4.9e-07;	
XX	Matches	61; Conservative 51; Mismatches 114; Indels 42; Gaps 8;	
QY	19	KGIAHIGYKALNELGIRVRLSGVSAIGALVSVFYASGYSPGPFSLKRVNWLKLPKFX 78	
DB	2	RGLYTAGVLDLADGAGIKIDGIVSVSAGALFGVNFVSRQERAL-----RYN--KYLISH 54	
QY	79	PLKELIGWEXAIRPLEEVLPRRI-----EKLEPTVICATDLYSGRALS 126	
DB	55	PKYMLGRNFRIGFVNRKDFYIEVPMKLDVDFDEAFKKSIDFTVVATEMTSGRPYFK 114	
QY	127	EGSL-----IPALLGSCAIPGIFEPVEYKYNLLVDGGIVNNLVPFQESGI-PTVCVDVLP 182	
DB	115	IDSVEPEQMEILRASSALPVVSKVMDQKKYLDGSLSDIPVDFAFGGLGFKLIVWTRP 174	
QY	183	TEPEKDIKN--ILHILLRSPFLAVRSNKKRKEFCF-LVIVPELEEFTEPL----- 229	
DB	175	LNQYKFSGGRLYKTLRYKYPNFVKYASNRYQYNNLSLEKWSLSEKTDGLFAIRPSKSLV 234	
QY	230	-----DVRKADQIMERGYIKALEVLSE 251	
DB	235	IGRLKPNPKLDSIYQLGMDAKSVMP 262	
XX	RESULT 15		
XX	AM25456		
ID	AM25456 standard; Protein; 282 AA.		
AC	AM25456;		
DT	16-OCT-2001 (first entry)		
XX	Human protein sequence SEQ ID NO:971.		
XX	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;		
KW	antifungal; anti-infective; anti-infective; anti-infective; anti-infective;		
KW	antibacterial; endocrine; cardiant; central nervous system; virucide;		
KW	anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;		
KW	antiallergic; haemostatic; vulnary; antidiabetic; osteopathic; eczema;		
KW	dermatological; antiallergic; antianaemic; antidiabetic; cytostatic;		
KW	neuroprotective; neurotropic; neurotropic; antiparkinsonian; infection;		
KW	immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;		
KW	antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;		
KW	cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;		
KW	genetic disease; haematopoietic disorder; platelet disorder; asthma;		
KW	thrombocytopenia; osteoporosis; severe combined immunodeficiency;		
KW	allergic rhinitis; diabetes; multiple sclerosis; depression;		
KW	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;		
XX	neurological disorder.		
OS	Homo sapiens.		
XX	WO200153455-A2.		
XX	26-JUL-2001.		
XX	22-DEC-2000; 2000WO-US35017.		
XX	23-DEC-1999; 99US-0471275.		
XX	21-JAN-2000; 2000US-0488725.		
XX	25-APR-2000; 2000US-0552317.		
XX	(HYSE-) HYSEQ INC.		
XX	Tang YT, Liu C, Drmanac RT;		
XX	WPI; 2001-457603/49.		
XX	N-PSDB; AAH99397.		
PT	Isolated human polynucleotides encoding polypeptides, useful for the		

PT	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -		
XX	Claim 20; Page 205; 1217pp; English.		
CC	AAH99166 to AAH99904 encode the human proteins given in AAH25225 to		
CC	AAH25963. The proteins can have activities based on the tissues and		
CC	cells they are expressed in, such as: anti-inflammatory; antineumatic;		
CC	antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;		
CC	central nervous system; virucide; anti-HIV; fungicide; antimutagen;		
CC	cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;		
CC	antidiabetic; osteopathic; dermatological; antiallergic; antisthmatic;		
CC	antiparkinsonian; and immunostimulant. The proteins and polynucleotides		
CC	encoding them can be used in gene therapy, antitense therapy and vaccine		
CC	production. The proteins and polynucleotides are useful for screening for		
CC	agonists or antagonists of a protein and for the treatment and diagnosis		
CC	of disorders associated with the activity of a protein e.g. inflammation,		
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,		
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal		
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,		
CC	anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,		
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic		
CC	rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,		
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and		
CC	neurological disorders.		
XX	Sequence 282 AA;		
SQ	Query Match	10.7%; Score 136; DB 22; Length 282;	
	Best Local Similarity	23.0%; Fred. No. 2.4e-06;	
	Matches	49; Conservative 37; Mismatches 91; Indels 36; Gaps 6;	
QY	60	EGMPSLLKRVNWLKFKPKPLKGLIGWEKAIRFLEEVLPYRRIEKLPIPTICATDLYS 119	
DB	17	EGMTSLMKAA-----LDLTYPITSMFSGAGFNSSIFSFKDQIEDLWIPYFAITDTA 71	
QY	120	GRALYLSGSLIPALLGSCAIPGIFEPV--EYKNVLLVDGGIVNNLVPFQESGIPTV- 176	
DB	72	SAMRVHTDGLNRYVRASMSLSGYNPPCLDPKDGHLMDGGYNNLPADVARSAGAKVI 131	
QY	177	CVDV-----LPIPEKDKNIIHILLRSFFFLAVRSNSE-- 209	
DB	132	AIDVGSRDETDLTNYGDALSGWMLLWKNWNLATKRVNMAEIQTRLAYVCCVRQLEV 191	
QY	210	KRKEFCDLVIVPELEEFTEPLDVRKADQIMERGY 242	
DB	192	KSSDYCF-YLRPPIDSVSYLDKCKFNEICEVGY 223	
	Search completed: June 19, 2003, 15:02:22		
	Job time : 45 secs		

Thu Jun 19 17:30:42 2003

us-09-903-410-36.rat

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 15:01:30 ; Search time 23 Seconds

(without alignments)  
321.094 Million cell updates/sec

Title: US-09-903-410-36

Sequence: 1 LRKRKEEINVLVSGAAKG.....RKADQIMRGVYIKALVLSH 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*

2: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*

3: /cgn2\_6/prodata/1/1aa/6A.COMB.pep:\*

4: /cgn2\_6/prodata/1/1aa/6B.COMB.pep:\*

5: /cgn2\_6/prodata/1/1aa/ECTUS.COMB.pep:\*

6: /cgn2\_6/prodata/1/1aa/Backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1265	99.4	251	2	US-08-602-359A-36
2	901.5	70.8	249	2	US-08-602-359A-39
3	86.5	6.8	1112	4	US-09-353-585-2
4	84.5	6.6	1112	4	US-09-353-585-3
5	81.5	6.4	393	2	US-08-977-554-2
6	81.5	6.4	393	2	US-09-225-967-2
7	81.5	6.4	393	4	US-09-227-806-2
8	81	6.4	832	1	US-07-977-434-12
9	81	6.4	832	1	US-08-458-819-12
10	81	6.4	832	5	PCT-US91-07035-12
11	78	6.1	488	4	US-09-099-053-2
12	77.5	6.1	408	3	US-09-074-912-2
13	77.5	6.1	408	3	US-09-074-912-4
14	77.5	6.1	408	3	US-09-290-136-2
15	77.5	6.1	408	4	US-09-290-136-4
16	76.5	6.0	984	2	US-08-673-789-6
17	76	6.0	350	2	US-08-852-481-2
18	74.5	5.9	357	2	US-09-193-191-2
19	74.5	5.9	951	1	US-08-162-809-2
20	74	5.8	444	4	US-09-413-814-95
21	74	5.8	444	4	US-09-413-814-96
22	74	5.8	537	4	US-09-134-001C-4091
23	73	5.7	545	4	US-08-687-590-26
24	73	5.7	843	4	US-09-361-631-2
25	73	5.7	3782	4	US-09-105-537-4
26	72.5	5.7	605	1	US-08-485-718-9
27	72.5	5.7	605	2	US-08-484-530-55

28	72.5	5.7	605	2	US-08-827-618A-55	Sequence 55, Appl
29	72.5	5.7	605	3	US-08-483-952A-55	Sequence 55, Appl
30	72.5	5.7	605	4	US-08-476-501-55	Sequence 55, Appl
31	72.5	5.7	662	4	US-09-134-001C-4074	Sequence 4074, Ap
32	72	5.7	673	4	US-08-186-387-8	Sequence 2, Appl
33	72	5.7	718	3	US-09-090-808-2	Sequence 2, Appl
34	72	5.7	718	4	US-09-447-453-2	Sequence 2, Appl
35	72	5.7	740	4	US-09-323-872A-25	Sequence 25, Appl
36	72	5.7	949	4	US-09-196-387-10	Sequence 10, Appl
37	72	5.7	1327	4	US-09-196-387-2	Sequence 2, Appl
38	71.5	5.6	269	4	US-09-355-166-11	Sequence 11, Appl
39	70	5.5	451	4	US-09-412-102-8	Sequence 8, Appl
40	70	5.5	451	4	US-09-217-787-8	Sequence 8, Appl
41	69.5	5.5	397	3	US-08-135-782-2	Sequence 2, Appl
42	69.5	5.5	554	3	US-08-904-871-1	Sequence 2, Appl
43	69.5	5.5	831	2	US-08-677-734A-11	Sequence 11, Appl
44	69.5	5.5	831	4	US-09-097-053-11	Sequence 11, Appl
45	69	5.4	245	4	US-09-369-364A-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-08-602-359A-36  
Sequence 36, Application US/08602359A  
Patent No. 5942430  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFFA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SPANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,359A  
FILING DATE: February 16, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HALL, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELEPHONE: 619-678-5099  
TELEFAX: 619-678-5097  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 251 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-602-359A-36  
Query Match 99.4%, Score 1265, DB 2, Length 251;

Best Local Similarity 99.6%; Pred. No. 9.4e-136;  
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LRLKFEINLVLSGGAAGIAHIGVLAINEIGIRVRLSGVSGAIVSVFYASGYSPE 60  
DB 1 LRLKFEINLVLSGGAAGIAHIGVLAINEIGIRVRLSGVSGAIVSVFYASGYSPE 60  
QY 61 GMPSLLRVNVKLKFKPKPLKGLIGWEKAIKRLFEVLVYRRIEKLPTTYICATDLYSG 120  
DB 61 GMPSLLRVNVKLKFKPKPLKGLIGWEKAIKRLFEVLVYRRIEKLPTTYICATDLYSG 120  
QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKVNLLVDGGIVNNLVPVEPQSGIPTVCVDV 180  
DB 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKVNLLVDGGIVNNLVPVEPQSGIPTVCVDV 180  
QY 181 LPIDPEKDKNIILHILRSFPLAVRSNKRKEFCDLVIVPELEEFPLDVRKADQIMER 240  
DB 181 LPIDPEKDKNIILHILRSFPLAVRSNKRKEFCDLVIVPELEEFPLDVRKADQIMER 240  
QY 241 GYIKALEVLS 251  
DB 241 GYIKALEVLS 251

RESULT 2  
US-08-602-359A-39  
; Sequence 39, Application US/08602359A  
; Patent No. 5942430  
; GENERAL INFORMATION:  
; APPLICANT: ROBERTSON, Daniel E.  
; APPLICANT: MURPHY, Dennis  
; APPLICANT: REID, John  
; APPLICANT: MAFFIA, Anthony  
; APPLICANT: LINK, Steven  
; APPLICANT: SWANSON, Ronald V.  
; APPLICANT: WARREN, Patrick V.  
; APPLICANT: KOSMOTKA, Anna  
; TITLE OF INVENTION: ESTERASES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & RICHARDSON P.C.  
; STREET: 4225 EXECUTIVE SQUARE, STE 1400  
; CITY: LA JOLLA  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/602,359A  
; FILING DATE: February 16, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAILE, LISA A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/010001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-678-5070  
; TELEFAX: 619-678-5099  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 AMINO ACIDS  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-602-359A-39

Query Match 70.8%; Score 901.5; DB 2; Length 249;  
Best Local Similarity 71.5%; Pred No. 2.1e-94;  
Matches 176; Conservative 31; Mismatches 38; Indels 1; Gaps 1;

QY 1 LRLKFEINLVLSGGAAGIAHIGVLAINEIGIRVRLSGVSGAIVSVFYASGYSPE 60  
DB 5 LKLRFEINLVLSGGAAGIAHIGVLAINEIGIRVRLSGVSGAIVSVFYASGYSPE 64  
QY 61 GMPSLLRVNVKLKFKPKPLKGLIGWEKAIKRLFEVLVYRRIEKLPTTYICATDLYSG 120  
DB 65 ENLKLLKEVNLKLFKFTPKMGLANGWEKAELEKELGVKELEDLAIPTYLCSADLYTG 124  
QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKVNLLVDGGIVNNLVPVEPQSGIPTVCVDV 180  
DB 125 RALYFGRGDLIPVLLGSKSIPGIFEPVEYENFLVDGGIVNNLVPVEPQSGIPTVCVDV 184  
QY 181 LPIDPEKDKNIILHILRSFPLAVRSNKRKEFCDLVIVPELEEFPLDVRKADQIMER 240  
DB 185 LPIDPEKDKNIILHILRSFPLAVRSNKRKEFCDLVIVPELEEFPLDVRKADQIMER 243  
QY 241 GYIKAL 246  
DB 244 GDMRAL 249

RESULT 3  
US-09-353-585-2  
; Sequence 2, Application US/09353585  
; Patent No. 6287665  
; GENERAL INFORMATION:  
; APPLICANT: Dixon, Mark S  
; APPLICANT: Jones, David A  
; APPLICANT: Jones, Jonathan DG  
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderhye PC  
; STREET: 8th Floor, 1100 No. 6287665th Glebe Road  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/353,585  
; FILING DATE: 15-Jul-1999  
; CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/930,277  
; FILING DATE: 27-OCT-1997  
; APPLICATION NUMBER: PCT/GB96/00785  
; FILING DATE: 01-APR-1996  
; APPLICATION NUMBER: GB 9506658.5  
; FILING DATE: 31-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ms Mary J Wilson  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 620-69  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1112 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <unknown>  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Tomato  
STRAIN: Cf2  
SEQUENCE DESCRIPTION: SEQ ID NO: 2  
US-09-353-585-2

Query Match 6.8%; Score 86.5; DB 4; Length 1112;  
Best Local Similarity 30.3%; Pred. No. 1.2;  
Matches 46; Conservative 25; Mismatches 56; Indels 25; Gaps 10;

QY 3 LRKEEINLV---ISGGAAGIAHIGVKAINEIGIRVALSG---VSAGAI--VSVPY- 53  
DB LNNLSRLNLTNNQSGSIPE---EIGYLRSLNDLGLSENALNGSI PASIGNNLSMVL 606  
QY 54 ---ASGYSPGMSILKRVNMLKFKRPPLKGLIGWEKA-IRPLEVLPYRRIEKL 108  
DB 607 YNNQSGSIPEEI-GYLSLTYLSL--GNNSLNGLI PASFAMRNQLALINDNNLIGE 663  
QY 109 PTTCATDLYSGRALYISGSL---IPALGS 137  
DB 664 PSSVC--NLTSLVLYMPRNMLKGVPCQCGN 693

## RESULT 4

US-09-353-585-3  
Sequence 3, Application US/09353585  
Patent No. 6287865

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
Jones, David A  
Jones, Jonathan DG  
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

## NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhuy PC  
STREET: 8th Floor, 1100 No. 6287865th Glebe Road  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22201-4714

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/353,585  
FILING DATE: 15-Jul-1999  
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/930,277  
FILING DATE: 27-OCT-1997  
APPLICATION NUMBER: PCT/GB96/00785  
FILING DATE: 01-APR-1996  
APPLICATION NUMBER: GB 950658.5  
FILING DATE: 31-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ms Mary J Wilson  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-69  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1112 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Tomato  
STRAIN: Cf2  
SEQUENCE DESCRIPTION: SEQ ID NO: 3  
US-09-353-585-3

Query Match 6.6%; Score 84.5; DB 4; Length 1112;  
Best Local Similarity 27.9%; Pred. No. 2;  
Matches 46; Conservative 23; Mismatches 45; Indels 51; Gaps 12;

QY 3 LRKEEINLV---ISGGAAGIAHIGVKAINEIGIRVALSG---VSAGAI--VSVPY- 53  
DB LNNLSRLNLTNNQSGSIPE---EIGYLRSLNDLGLSENALNGSI PASIGNNLSMVL 606  
QY 54 ---ASGYSPGMSILKRVNMLKFKRPPLKGLIGWEKAIRPLEVLP---YRRI 104  
DB 607 YNNQSGSIPEEI-GYLSLTYLSL--GNNSLNGLI PASFAMRNQLALINDNNLIGE 650  
QY 105 KL-----EPTTCATDLYSGRALYISGSL---IPALGS 137  
DB 651 ALINDNNLIGRIPSSVC--NLTSLVLYMPRNMLKGVPCQCGN 693

## RESULT 5

US-08-977-554-2  
Sequence 2, Application US/08977554  
Patent No. 5891672

## GENERAL INFORMATION:

APPLICANT: Palmer, Leslie M.  
APPLICANT: Pedon, Jason C.  
APPLICANT: Warren, Richard L.  
APPLICANT: Train, Christopher M.  
APPLICANT: Wang, Min  
APPLICANT: Jaworski, Deborah D.  
APPLICANT: Mooney, Jeffrey  
APPLICANT: Debonck, Christine  
APPLICANT: Zhong, Yiyi  
APPLICANT: Black, Michael

## TITLE OF INVENTION: r1ba

NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/977,554  
FILING DATE:

## CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/02318  
FILING DATE: 19-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, O. Todd  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50444-07  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215/994-2252  
TELEFAX: 215/994-2222  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-977-554-2

Query Match 6.4%; Score 81.5; DB 2; Length 393;  
Best Local Similarity 22.1%; Pred. No. 0.88;  
Matches 58; Conservative 39; Mismatches 91; Indels 75; Gaps 13;  
QY 32 ELGIRVALSGVS-AGAIVSVFYASGSPGMSLLKRVNWLKFKPKPLKGLIGWEKA 90  
DB 144 EAAVDLAKLTGAKPAGVCEIMNDGDTMAKGOD-----LQNFKEKHOLKMIT----- 190  
QY 91 IRFLEEVLPYRR-----IE---KLEIPYICATDLYSGRALYLSGSLI----- 131  
DB 191 ---IDDLIEYRKLEPEIEFKAKVMPDTFGTDMYGFKATYDEEIVLTGKAIRQHEN 247  
QY 132 PALLGSCAIPGIF-----EPVEYKN-----YLLVDG---GIVNNLPVEPFQ 169  
DB 248 VRLHSACTGDTGFHSORCDGAGLESSMKYINEHGMIYLPQEGRGIGLNLKRAYELI 307  
QY 218 VIVPELEBFTPLDVRKADQIMER 240  
DB 363 DIAERIEVIVPETVHNHDYMTK 385

## RESULT 6

US-09-225-967-2  
; Sequence 2, Application US/09225967  
; Patent No. 6171598  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Leslie M.  
; APPLICANT: Fedon, Jason C.  
; APPLICANT: Warren, Richard L.  
; APPLICANT: Traini, Christopher M.  
; APPLICANT: Wang, Min  
; APPLICANT: Jaworski, Deborah D.  
; APPLICANT: Mooney, Jeffrey  
; APPLICANT: Debouck, Christine  
; APPLICANT: Zhong, Yiyi  
; APPLICANT: Black, Michael  
; TITLE OF INVENTION: riba  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; APPLICATION NUMBER: US/09/225,967  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/977,554  
; FILING DATE:  
; APPLICATION NUMBER: PCT/US97/02318  
; FILING DATE: 19-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Q. Todd  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER: P50444-07  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215/994-2252

TELEFAX: 215/994-2222  
TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 393 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-225-967-2

Query Match 6.4%; Score 81.5; DB 4; Length 393;  
Best Local Similarity 22.1%; Pred. No. 0.86;  
Matches 59; Conservative 39; Mismatches 91; Indels 75; Gaps 13;  
QY 32 ELGIRVALSGVS-AGAIVSVFYASGSPGMSLLKRVNWLKFKPKPLKGLIGWEKA 90  
DB 144 EAAVDLAKLTGAKPAGVCEIMNDGDTMAKGOD-----LQNFKEKHOLKMIT----- 190  
QY 91 IRFLEEVLPYRR-----IE---KLEIPYICATDLYSGRALYLSGSLI----- 131  
DB 191 ---IDDLIEYRKLEPEIEFKAKVMPDTFGTDMYGFKATYDEEIVLTGKAIRQHEN 247  
QY 132 PALLGSCAIPGIF-----EPVEYKN-----YLLVDG---GIVNNLPVEPFQ 169  
DB 248 VRLHSACTGDTGFHSORCDGAGLESSMKYINEHGMIYLPQEGRGIGLNLKRAYELI 307  
QY 170 ESGIPTVCVDVLPFIEPEKDIK-----NILHILLSFPLAVRSNSEKKEPCDL 217  
DB 308 EOGYDTVTAN-LAIGFDEDLRDYHIAAQILKYFNIHINLLS---NNPSKPEGLKQY-GI 362  
QY 218 VIVPELEBFTPLDVRKADQIMER 240  
DB 363 DIAERIEVIVPETVHNHDYMTK 385

## RESULT 7

US-09-227-806-2  
; Sequence 2, Application US/09227806  
; Patent No. 6280971  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Leslie M.  
; APPLICANT: Fedon, Jason C.  
; APPLICANT: Warren, Richard L.  
; APPLICANT: Traini, Christopher M.  
; APPLICANT: Wang, Min  
; APPLICANT: Jaworski, Deborah D.  
; APPLICANT: Mooney, Jeffrey  
; APPLICANT: Debouck, Christine  
; APPLICANT: Zhong, Yiyi  
; APPLICANT: Black, Michael  
; TITLE OF INVENTION: riba  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; APPLICATION NUMBER: US/09/227,806  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/977,554  
; FILING DATE:  
; APPLICATION NUMBER: PCT/US97/02318  
; FILING DATE: 19-FEB-1997



ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, O. Todd  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50444-07  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215/994-2252  
TELEFAX: 215/994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-227-806-2

Query Match 6.4%; Score 81.5; DB 4; Length 393;  
Best Local Similarity 22.1%; Pred. No. 0.88; Mismatches 91; Indels 75; Gaps 13;  
Matches 58; Conservative 39;

QY 32 ELGIRVALSGVS-AGAIIVSVPYASGYSPEGMFSLKRYNMKLFKFKPLKGLIGWEKA 90  
DB 144 EAAVDLAKLTGAKPAGVTCIEMDDGTMAKGD-----LQNKKEKGLKMIT----- 190  
QY 91 IRPEEVLPRR-----IE---KLEPTTCATDLYSGRALYLSGSLI----- 131  
DB 191 ---IDDLIERKLEPEIEFKAKVKMPTGFTDMYGKATYTBELIVLTGKAIROHEN 247  
QY 132 PALGSCAIPGIF-----EPVEYKN-----YLLVGD-----GIYNNLPVEFQ 169  
DB 248 VRLHSACLTGDIIFISORCDCAQLSESMKMYINHGAMITYLPQERGGIGLTKRAVELI 307  
QY 170 ESGIPTCVGVLPPEPEKDK-----NIHLILSFFLAIVNSSEKKEKPCDL 217  
DB 308 EGGIDYTAN-LALGPEDEADRDYHIAQLTKYFNLEHNLIS--NNSKEKGLKQY-GI 362  
QY 218 VIVELEBFTPLDVRKADQIMER 240  
DB 363 DIARIEVIVPEYVHNHDVMTK 385

RESULT 8  
US-07-977-434-12  
Sequence 12, Application US/07977434  
Patent No. 5466591.  
GENERAL INFORMATION:  
APPLICANT: Gelfand, David H.  
APPLICANT: Abramson, Richard D.  
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
ZIP: 07110-1199  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 7  
SOFTWARE: WordPerfect 2.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/977,434  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,490  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,466  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 590,213  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 523,394  
FILING DATE: 15-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 143,441  
FILING DATE: 12-JAN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 063,509  
FILING DATE: 17-JUN-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,241  
FILING DATE: 22-AUG-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 746,121  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US90/07641  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 585,471  
FILING DATE: 20-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 455,611  
FILING DATE: 22-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 609,157  
FILING DATE: 02-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 557,517  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Iuann Casert  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2972  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 892 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-977-434-12

Query Match 6.4%; Score 81; DB 1; Length 892;  
Best Local Similarity 22.6%; Pred. No. 3.5;  
Matches 53; Conservative 47; Mismatches 75; Indels 60; Gaps 14;

QY 26 VLAINEIGIRVALSGVSAGAIIVSVPYASGYSPEGMFSLKRYNMKLFKFKPLKGLI 65  
DB 95 VEEVLDLGIKTKIGFADDIITL-----SKKEESDEKKNV-----IT 136  
QY 86 GMEKAIPEEVLPRRIKLEIPYICATDLYSGRALYLSGSLIPALGSCAIPGIFE 145  
DB 137 GDXDLQLVSDKVFVRVR-----GITDL-----VLV-DRNKVIEKY-----IT 176  
QY 146 PVEKQVILVDDGIIVNLVPEPFGSGIPTV-----CYDVLPIPEKDKIKNL---HILR 198  
DB 177 PEQFKDILSVGDQIDNP-----GVKGIQKTNVSL--KKNVLEAVLKNINILIR 227  
QY 199 SFFLAIVNSSEKKEKPCDLVIVPELEBFTPLDVRKADQIMERG--IKALEVISE 251  
DB 228 KLRRLDESKEDLOKSIELV---ELIYDPMVDER-DEIIRGYNDPKLKLKVLK 278

RESULT 9  
US-08-458-819-12  
Sequence 12, Application US/08458819  
Patent No. 5795762  
GENERAL INFORMATION:  
APPLICANT: Gelfand, David H.



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 585,471  
FILING DATE: 20-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 455,611  
FILING DATE: 22-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 609,157  
FILING DATE: 02-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 557,517  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SLAG PH.D. STACEY R.  
REGISTRATION NUMBER: 32,630  
REFERENCE/DOCKET NUMBER: Case No. 2580  
TELEPHONE: 415-420-3300  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 892 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-07035-12

Query Match 6.4%; Score 81; DB 5; Length 892;  
Best Local Similarity 22.6%; Pred. No. 3.5;  
Matches 53; Conservative 47; Mismatches 75; Indels 60; Gaps 14;

QY 26 VLKAINELGIRVRLSGVSAGAIVSFVSGVSPGMSLKRNVMLKLFKPKLKGIL 85  
DB 95 VEEIVDNLGKIKVKBIFRADIATL-----SKKFSDEPKNI-----if 136  
QY 86 GMEKAIKRLBEVLPRRIEGLIPTCATDYSGRALYSEGLIPALGSCAIPGIFE 145  
DB 137 GDKDLQVSDKVPVWVWER-----GITDL-----VLY-DRNKVIKRY-----GIY- 176  
QY 146 PVEYKNYLLVDGIVNNLPVEPQESGIPTV-----CDVYLPIEPEKQIKNLL--HILR 198  
DB 177 PEGFKDYLISLVDQIDNIF-----GVGICKKTRAVSL--KKNSLENNLKNINLITE 227  
QY 199 SFLAVNSNKRKFCDLVTVPELEFPLVYKRAQQIMERG--IKALEVLSB 251  
DB 228 KLRRLBDSKEDLQKSIETV--ELIYDVPMVER-DELIYGVNPDLLKVLKK 278

## RESULT 11

US-09-099-053-2  
Sequence 2, Application US/09099053  
Patent No. 6388063  
GENERAL INFORMATION:  
APPLICANT: Greg Plowman  
APPLICANT: Susan Onrust  
APPLICANT: David Markby  
APPLICANT: Sara Courtneidge  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
NUMBER OF INVENTIONS: SAD RELATED DISORDERS  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Filth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,053  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/049,914  
FILING DATE: June 18, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 235/121  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 488 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-099-053-2

Query Match 6.1%; Score 78; DB 4; Length 488;  
Best Local Similarity 21.9%; Pred. No. 3.1;  
Matches 42; Conservative 27; Mismatches 75; Indels 48; Gaps 6;

QY 6 FEEINVLISGAKAKIAHIGV-----LKAINELG--IRVRLSGVAGAIIVGVFASGS 58  
DB 52 FPGFLATVDFARCGEISVRGDRLCALREGGYIPARISQGSAGLVITVAAS 111  
QY 59 PEM-----ESLIRVNMMLK-----PKRPLKGLGMEKAIKRLBEVLPRR 102  
DB 112 PETLSQPMYFEGVRRTOQQLSPNREAFIRSESSLGGYLSVRAQAKVCHR- 170  
QY 103 IETRLIPTICATDYSGRALYSEGLIPALGSCAIPGIFPVEYKNYLLVDGIVNN 162  
DB 171 -----VSMADGSLYLQGRLPGL-----EBLLTYKANWMLION 206  
QY 163 LPVEPQESGIP 174  
DB 207 PLIQPCMPQKAP 218

## RESULT 12

US-09-074-912-2  
Sequence 2, Application US/09074912  
Patent No. 6057491  
GENERAL INFORMATION:  
APPLICANT: CIGAN, AMY L.  
APPLICANT: CZAPLA, THOMAS H.  
APPLICANT: FALLIS, LYNN  
APPLICANT: MEYER, TERRY E.  
APPLICANT: MUNDELL, SCOTT A.  
APPLICANT: SABUS, BRIAN  
APPLICANT: SCHUBERT, KAREL  
TITLE OF INVENTION: PROTEINS HAVING INSECTICIDAL ACTIVITIES  
NUMBER OF INVENTIONS: AND METHODS OF USE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: W. MORRAY SPRUILL (ALSTON & BIRD, LLP)  
STREET: 3605 GLENWOOD AVE.  
CITY: RALEIGH  
STATE: NC  
COUNTRY: USA  
ZIP: 27622  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/074,912  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: SPRUILL, W. MURRAY  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: 5718-9  
TELEPHONE: 919 420 2202  
TELEFAX: 919 881 3175  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-074-912-2

Query Match  
Best Local Similarity 6.1%; Score 77.5; DB 3; Length 408;  
Matches 51; Conservative 34; Mismatches 77; Indels 71; Gaps 12;

QY 11 LVLGGGAAGIAHIGVLKAIN-----ELGIRVRALSGVSGAIVSVFYASG---- 56  
DB 39 LAIDGGGIRGIIPGVILKQLEATLQWDSARLAIEYFDVAGTSTGGIITAILTAPDPQN 98  
QY 57 -----YSPGGMF-----SLLKRVNMLKLFKFKPLKGLIGWEKAIRFLERVLPRR 102  
DB 99 KDRPLAAEIIIDFYIEHGPSIFNKSTACSLPGIFCPKYDG-----KYLQIIISQKL 150  
QY 103 IEKL-----EIPYICATDLYSGRALYLSGSL--IPAL-----LGSCAIPGIF 144  
DB 151 NETLLDQTTNNVIPSF-----DIKLLRPTIFSTFKLEVPPELVNKLSDVCMGTSAAPIVF 206  
QY 145 EPVEYKN-----YLLVDGGIVNNLPVEPFQESGIPTCVDVLPPIEPKDIKNIL 193  
DB 207 PPYFKHGDTFNLVDGAIADIPA-----PVALSEVLQOEKYKN-KEIL 250

RESULT 13  
US-09-074-912-4  
Sequence 4, Application US/09074912  
Patent No. 6057491  
GENERAL INFORMATION:  
APPLICANT: CIGAN, AMY L.  
APPLICANT: CZAPLA, THOMAS H.  
APPLICANT: FALLIS, LYNN  
APPLICANT: MEYER, TERRY E.  
APPLICANT: MONDELL, SCOTT A.  
APPLICANT: SABUS, BRIAN  
APPLICANT: SCHUBERT, KAREL  
TITLE OF INVENTION: PROTEINS HAVING INSECTICIDAL ACTIVITIES  
TITLE OF INVENTION: AND METHODS OF USE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: W. MURRAY SPRUILL (ALSTON & BIRD, LLP)  
STREET: 3605 GLENWOOD AVE.  
CITY: RALEIGH  
STATE: NC  
COUNTRY: USA  
ZIP: 27622  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,912  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: SPRUILL, W. MURRAY  
REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: 5718-9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919 420 2202  
TELEFAX: 919 881 3175  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-074-912-4

Query Match  
Best Local Similarity 6.1%; Score 77.5; DB 3; Length 408;  
Matches 51; Conservative 34; Mismatches 77; Indels 71; Gaps 12;

QY 11 LVLGGGAAGIAHIGVLKAIN-----ELGIRVRALSGVSGAIVSVFYASG---- 56  
DB 39 LAIDGGGIRGIIPGVILKQLEATLQWDSARLAIEYFDVAGTSTGGIITAILTAPDPQN 98  
QY 57 -----YSPGGMF-----SLLKRVNMLKLFKFKPLKGLIGWEKAIRFLERVLPRR 102  
DB 99 KDRPLAAEIIIDFYIEHGPSIFNKSTACSLPGIFCPKYDG-----KYLQIIISQKL 150  
QY 103 IEKL-----EIPYICATDLYSGRALYLSGSL--IPAL-----LGSCAIPGIF 144  
DB 151 NETLLDQTTNNVIPSF-----DIKLLRPTIFSTFKLEVPPELVNKLSDVCMGTSAAPIVF 206  
QY 145 EPVEYKN-----YLLVDGGIVNNLPVEPFQESGIPTCVDVLPPIEPKDIKNIL 193  
DB 207 PPYFKHGDTFNLVDGAIADIPA-----PVALSEVLQOEKYKN-KEIL 250

RESULT 14  
US-09-290-136-2  
Sequence 2, Application US/09290136  
Patent No. 6339144  
GENERAL INFORMATION:  
APPLICANT: Cigan, Amy L.  
APPLICANT: Czapl, Thomas H.  
APPLICANT: Fallis, Lynn  
APPLICANT: Meyer, Terry E.  
APPLICANT: Mundell, Scott A.  
APPLICANT: Sabus, Brian  
APPLICANT: Schubert, Karel  
TITLE OF INVENTION: Proteins Having Insecticidal Activities and Methods of  
TITLE OF INVENTION: Use  
FILE REFERENCE: 5718-9A, 035718/180486  
CURRENT APPLICATION NUMBER: US/09/290,136  
EARLIER FILING DATE: 1999-04-13  
EARLIER FILING DATE: 09/074,912  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 408  
TYPE: PRT  
ORGANISM: Pentaclethra macroloba  
US-09-290-136-2

Query Match  
Best Local Similarity 6.1%; Score 77.5; DB 4; Length 408;  
Matches 51; Conservative 34; Mismatches 77; Indels 71; Gaps 12;

QY 11 LVLGGGAAGIAHIGVLKAIN-----ELGIRVRALSGVSGAIVSVFYASG---- 56  
DB 39 LAIDGGGIRGIIPGVILKQLEATLQWDSARLAIEYFDVAGTSTGGIITAILTAPDPQN 98  
QY 57 -----YSPGGMF-----SLLKRVNMLKLFKFKPLKGLIGWEKAIRFLERVLPRR 102  
DB 99 KDRPLAAEIIIDFYIEHGPSIFNKSTACSLPGIFCPKYDG-----KYLQIIISQKL 150  
QY 103 IEKL-----EIPYICATDLYSGRALYLSGSL--IPAL-----LGSCAIPGIF 144

Thu Jun 19 17:30:42 2003

us-09-903-410-36.ra1

Page 9

DB 151 NETLDDQTTNVPISF---DILKLRPTISTFKLEVEPELNVKLSDVCMGTSAPYIF 206  
QY 145 EPEVEYKN---YLLVDGGIIVNNLPEPPEQESGIPVVCVDVLPLEPEKDIKNIL 193  
DB 207 PPEYFKHGDTEFENLVDAIADIIPA-----FVALSEVLQOEKYN-KEIL 250

RESULT 15

US-09-290-136-4  
; Sequence 4, Application US/09290136  
; Patent No. 6339144  
; GENERAL INFORMATION:  
; APPLICANT: Cigan, Amy L.  
; APPLICANT: Czaplak, Thomas H.  
; APPLICANT: Fallis, Lynn  
; APPLICANT: Meyer, Terry E.  
; APPLICANT: Mundell, Scott A.  
; APPLICANT: Sabus, Brian  
; APPLICANT: Schubert, Karel  
; TITLE OF INVENTION: Proteins Having Insecticidal Activities and Methods of  
; TITLE OF INVENTION: Use  
; FILE REFERENCE: 5718-9A, 035718/180486  
; CURRENT APPLICATION NUMBER: US/09/290,136  
; CURRENT FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 09/074,912  
; EARLIER FILING DATE: 1998-05-08  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 408  
; TYPE: PRT  
; ORGANISM: Pentaclethra macroloba  
US-09-290-136-4

Query Match 6.1%; Score 77.5; DB 4; Length 408;

Best Local Similarity 21.9%; Pred. No. 2.6; Matches 51; Conservative 34; Mismatches 77; Indels 71; Gaps 12;

DB 11 LVLGGAAKGINHIGVLRKAIN-----ELGIRVRALSGVAGAIIVSVFYASG----- 56  
QY 39 LAIDGGGIRGILPGVILKQLENTLQRMWSSARLAEYFDVAVAGTSTGGIITAILTADPQN 98  
DB 57 -----YSPRGMF-----SLKRVNMLKU-FKPKPLKGLIGWEKAIKFLBEVLPYRR 102  
QY 99 KDRPLVAAEELIDFYIEHGPSIFPNKSTACSLPGIFCPKPDG-----KYLQELISOKL 150  
DB 103 IEKL-----ELPFIYICATDLYSGRALYSEGSU--IPAL-----LGSCAIPGIF 144  
QY 151 NETLDDQTTNVPISF---DILKLRPTISTFKLEVEPELNVKLSDVCMGTSAPYIF 206  
DB 145 EPEVEYKN---YLLVDGGIIVNNLPEPPEQESGIPVVCVDVLPLEPEKDIKNIL 193  
QY 207 PPEYFKHGDTEFENLVDAIADIIPA-----FVALSEVLQOEKYN-KEIL 250

Search completed: June 19, 2003, 15:05:27  
Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 19, 2003, 15:04:11 ; Search time 182 Seconds

(without alignments)  
149.230 Million cell updates/sec

Title: US-09-903-410-36

Sequence: 1273  
1 LRLKREINIVLSGGAAGK.....RKADQIMRGYKALEVLSR 251

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 10820613 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEM\_PUB pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1273	100.0	251	10	US-09-903-410-36 Sequence 36, Appl
2	1265	99.4	251	9	US-10-027-805-36 Sequence 36, Appl
3	1265	99.4	251	9	US-10-027-804-36 Sequence 36, Appl
4	913.5	71.8	249	10	US-09-903-410-39 Sequence 39, Appl
5	901.5	70.8	249	9	US-10-027-805-39 Sequence 39, Appl
6	901.5	70.8	249	9	US-10-027-804-39 Sequence 39, Appl
7	139.5	11.0	425	9	US-10-147-026-10 Sequence 10, Appl
8	109.5	8.6	504	10	US-09-765-205-10 Sequence 40, Appl
9	86.5	6.8	390	9	US-10-034-937-14 Sequence 14, Appl
10	85.5	6.7	3354	9	US-10-160-758-11 Sequence 11, Appl
11	85.5	6.7	3354	9	US-10-160-758-12 Sequence 12, Appl
12	83.5	6.6	390	9	US-10-034-937-10 Sequence 10, Appl
13	83.5	6.6	455	9	US-10-082-627A-25 Sequence 25, Appl
14	83	6.5	391	10	US-09-815-242-5379 Sequence 5379, A
15	83	6.5	393	10	US-09-815-242-12285 Sequence 12285, A
16	81.5	6.4	390	9	US-10-034-937-12 Sequence 12, Appl
17	81.5	6.4	390	9	US-10-034-937-12 Sequence 12, Appl
18	81	6.4	802	9	US-10-245-103-78 Sequence 78, Appl
19	81	6.4	802	9	US-10-245-107-78 Sequence 78, Appl

20	81	6.4	802	9	US-10-245-143-78	Sequence 78, Appl
21	81	6.4	802	9	US-10-245-771-78	Sequence 78, Appl
22	81	6.4	802	9	US-10-245-851-78	Sequence 78, Appl
23	81	6.4	802	9	US-10-245-883-78	Sequence 78, Appl
24	81	6.4	802	9	US-10-237-535-78	Sequence 78, Appl
25	81	6.4	802	9	US-10-338-183-78	Sequence 78, Appl
26	81	6.4	802	9	US-10-338-283-78	Sequence 78, Appl
27	81	6.4	802	9	US-10-338-370-78	Sequence 78, Appl
28	81	6.4	802	9	US-10-245-055-78	Sequence 78, Appl
29	81	6.4	802	9	US-10-245-147-78	Sequence 78, Appl
30	81	6.4	802	9	US-10-245-730-78	Sequence 78, Appl
31	81	6.4	802	9	US-10-245-739-78	Sequence 78, Appl
32	81	6.4	802	9	US-10-246-210-78	Sequence 78, Appl
33	81	6.4	802	9	US-10-239-196-78	Sequence 78, Appl
34	81	6.4	802	9	US-10-243-024-78	Sequence 78, Appl
35	81	6.4	802	9	US-10-243-409-78	Sequence 78, Appl
36	81	6.4	802	9	US-10-245-033-78	Sequence 78, Appl
37	81	6.4	802	9	US-10-245-621-78	Sequence 78, Appl
38	81	6.4	802	9	US-10-245-880-78	Sequence 78, Appl
39	81	6.4	802	9	US-10-243-095-78	Sequence 78, Appl
40	81	6.4	802	9	US-10-245-185-78	Sequence 78, Appl
41	81	6.4	802	9	US-10-245-427-78	Sequence 78, Appl
42	81	6.4	802	9	US-10-245-473-78	Sequence 78, Appl
43	81	6.4	802	9	US-10-245-770-78	Sequence 78, Appl
44	81	6.4	802	9	US-10-245-877-78	Sequence 78, Appl
45	81	6.4	802	9	US-10-246-976-78	Sequence 78, Appl

## ALIGNMENTS

RESULT 1  
US-09-903-410-36  
Sequence 36, Application US/09903410  
Patent No. US20020146799A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
APPLICANT: ROBERTSON, Dan  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MARFIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald  
APPLICANT: WARREN, Patrick  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ENZYMS HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF  
FILE REFERENCE: DIVER1180-2  
CURRENT APPLICATION NUMBER: US/09/903, 410  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 09/382,242  
PRIOR FILING DATE: 1999-08-24  
PRIOR APPLICATION NUMBER: US 08/602,359  
PRIOR FILING DATE: 1996-02-16  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 36  
LENGTH: 251  
TYPE: PRT  
ORGANISM: Aquifex pyrophilus  
US-09-903-410-36

Query Match 100.0%; Score 1273; DB 10; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.2e-120;  
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	LRLKREINIVLSGGAAGKIAITGKAINELGIVRLSLGVSAGAIYSVYAGSYSE 60
DB	1	LRLKREINIVLSGGAAGKIAITGKAINELGIVRLSLGVSAGAIYSVYAGSYSE 60
QY	61	GMFSLKRVVWLKLFKFKPPKLGKILGMEKAIKRLBEVLPYRRIEKLIEPTTICATDIVSG 120
DB	61	GMFSLKRVVWLKLFKFKPPKLGKILGMEKAIKRLBEVLPYRRIEKLIEPTTICATDIVSG 120

Qy 121 RALYLSGLIPALLGSCAIGIPEPVEYKYNLLVDGGIIVNNLPVPPFOESGIPVTCVDV 180  
 Db 121 RALYLSGLIPALLGSCAIGIPEPVEYKYNLLVDGGIIVNNLPVPPFOESGIPVTCVDV 180  
 Qy 181 LPTEPEKDKNHLHLLRSFFLAVERNSSEKKEFCDLVIVPELEETPLDVRKADQIMER 240  
 Db 181 LPTEPEKDKNHLHLLRSFFLAVERNSSEKKEFCDLVIVPELEETPLDVRKADQIMER 240  
 Qy 241 GYIKALEVLSE 251  
 Db 241 GYIKALEVLSE 251

RESULT 2

US-10-027-805-36  
 ; Sequence 36, Application US/10027805  
 ; Patent No. US20020164725A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROBERTSON, Daniel B.  
 ; MURPHY, Dennis  
 ; REID, John  
 ; MAFFIA, Anthony  
 ; LINK, Steven  
 ; SWANSON, Ronald V.  
 ; WARREN, Patrick V.  
 ; KOSMOTKA, Anna  
 ; TITLE OF INVENTION: ESTERASES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FISH & RICHARDSON P.C.  
 ; STREET: 4225 EXECUTIVE SQUARE, STE 1400  
 ; CITY: LA JOLLA  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 INCH DISKETTE  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WORD PERFECT 6.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/027.805  
 ; FILING DATE: 21-Dec-2001  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/602,359  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: HAILE, LISA A.  
 ; REFERENCE/DOCKET NUMBER: 09010/010001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-678-5070  
 ; TELEFAX: 619-678-5099  
 ; INFORMATION FOR SEQ ID NO: 36:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 251 AMINO ACIDS  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: PROTEIN  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 36:  
 US-10-027-805-36

Query Match 99.4%; Score 1265; DB 9; Length 251;  
 Best Local Similarity 99.6%; Pred. No. 7.6e-120;  
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 LRLRKEEINLVLSGGAAGIAHIGVLAINEIRVRLSGVSGAIVSVFYASGYSPE 60  
 Db 1 LRLRKEEINLVLSGGAAGIAHIGVLAINEIRVRLSGVSGAIVSVFYASGYSPE 60  
 Qy 61 GMFSLKRVNMLKFKPKPLKGLIGWEKAIRFLEEVLPYRRIEKLEIPTYICATDLYSG 120

Db 61 GMFSLKRVNMLKFKPKPLKGLIGWEKAIRFLEEVLPYRRIEKLEIPTYICATDLYSG 120  
 Qy 121 RALYLSGLIPALLGSCAIGIPEPVEYKYNLLVDGGIIVNNLPVPPFOESGIPVTCVDV 180  
 Db 121 RALYLSGLIPALLGSCAIGIPEPVEYKYNLLVDGGIIVNNLPVPPFOESGIPVTCVDV 180  
 Qy 181 LPTEPEKDKNHLHLLRSFFLAVERNSSEKKEFCDLVIVPELEETPLDVRKADQIMER 240  
 Db 181 LPTEPEKDKNHLHLLRSFFLAVERNSSEKKEFCDLVIVPELEETPLDVRKADQIMER 240  
 Qy 241 GYIKALEVLSE 251  
 Db 241 GYIKALEVLSE 251

RESULT 3

US-10-027-804-36  
 ; Sequence 36, Application US/10027804  
 ; Publication No. US20030054530A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROBERTSON, Daniel B.  
 ; MURPHY, Dennis  
 ; REID, John  
 ; MAFFIA, Anthony  
 ; LINK, Steven  
 ; SWANSON, Ronald V.  
 ; WARREN, Patrick V.  
 ; KOSMOTKA, Anna  
 ; TITLE OF INVENTION: ESTERASES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FISH & RICHARDSON P.C.  
 ; STREET: 4225 EXECUTIVE SQUARE, STE 1400  
 ; CITY: LA JOLLA  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 INCH DISKETTE  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WORD PERFECT 6.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/027.804  
 ; FILING DATE: 21-Dec-2001  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/602,359  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: HAILE, LISA A.  
 ; REFERENCE/DOCKET NUMBER: 09010/010001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-678-5070  
 ; TELEFAX: 619-678-5099  
 ; INFORMATION FOR SEQ ID NO: 36:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 251 AMINO ACIDS  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: PROTEIN  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 36:  
 US-10-027-804-36

Query Match 99.4%; Score 1265; DB 9; Length 251;  
 Best Local Similarity 99.6%; Pred. No. 7.6e-120;  
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 LRLRKEEINLVLSGGAAGIAHIGVLAINEIRVRLSGVSGAIVSVFYASGYSPE 60  
 Db 1 LRLRKEEINLVLSGGAAGIAHIGVLAINEIRVRLSGVSGAIVSVFYASGYSPE 60

QY 61 GMSFLKRVNMLKLFKFKPKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLYSG 120  
DB 61 GMSFLKRVNMLKLFKFKPKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLYSG 120  
QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVY 180  
DB 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVY 180  
QY 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDLYVPELBEFTPLDVRKADQIMER 240  
DB 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDLYVPELBEFTPLDVRKADQIMER 240  
QY 241 GYIKALEVLS 251  
DB 241 GYIKALEVLS 251

RESULT 4  
US-09-903-410-39  
Sequence 39, Application US/09903410  
Patent No. US20020146799A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
APPLICANT: ROBERTSON, Dan  
APPLICANT: MOREHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFITA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald  
APPLICANT: WARREN, Patrick  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ENZYMS HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF  
FILE REFERENCE: DIVER1180-2  
CURRENT FILING DATE: US/09/903,410  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 09/382,242  
PRIOR FILING DATE: 1999-08-24  
PRIOR APPLICATION NUMBER: US 08/602,359  
PRIOR FILING DATE: 1996-02-16  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 39  
LENGTH: 249  
TYPE: PRT  
ORGANISM: Aquifex VF5-34LC  
US-09-903-410-39

Query Match 71.8%; Score 913.5; DB 10; Length 249;  
Best Local Similarity 72.0%; Pred. No. 2.6e-84;  
Matches 177; Conservative 31; Mismatches 37; Indels 1; Gaps 1;  
QY 1 LRLKREERNLVLSSGAAGIAHIGVLAINELGIRVALSGVSAAGIYVFAASGSPR 60  
DB 5 LKLRKEEENLVLSSGAAGIAHIGVLAINELGIRVALSGVSAAGIYVFAASGSPR 64  
QY 61 GMSFLKRVNMLKLFKFKPKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLYSG 120  
DB 61 GMSFLKRVNMLKLFKFKPKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLYSG 124  
QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVY 180  
DB 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVY 184  
QY 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDLYVPELBEFTPLDVRKADQIMER 240  
DB 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDLYVPELBEFTPLDVRKADQIMER 243  
QY 241 GYIKAL 246  
DB 244 GDMRAL 249

RESULT 5

US-10-027-805-39  
Sequence 39, Application US/10027805  
Patent No. US20020164725A1  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MOREHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFITA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION NUMBER: US/10/027,805  
FILING DATE: 21-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,359  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HALE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 249 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
US-10-027-805-39

Query Match 70.8%; Score 901.5; DB 9; Length 249;  
Best Local Similarity 71.5%; Pred. No. 4.3e-83;  
Matches 176; Conservative 31; Mismatches 38; Indels 1; Gaps 1;  
QY 1 LRLKREERNLVLSSGAAGIAHIGVLAINELGIRVALSGVSAAGIYVFAASGSPR 60  
DB 5 LKLRKEEENLVLSSGAAGIAHIGVLAINELGIRVALSGVSAAGIYVFAASGSPR 64  
QY 61 GMSFLKRVNMLKLFKFKPKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLYSG 120  
DB 61 GMSFLKRVNMLKLFKFKPKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLYSG 124  
QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVY 180  
DB 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVY 184  
QY 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDLYVPELBEFTPLDVRKADQIMER 240  
DB 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDLYVPELBEFTPLDVRKADQIMER 243  
QY 241 GYIKAL 246  
DB 244 GDMRAL 249



RESULT 6  
 US-10-027-804-39  
 ; Sequence 39, Application US/10027804  
 ; Publication No. US20030054530A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROBERTSON, Daniel E.  
 ; MURPHY, Dennis  
 ; REID, John  
 ; MAFIA, Anthony  
 ; LINK, Steven  
 ; SWANSON, Ronald V.  
 ; WARREN, Patrick V.  
 ; KOSMOTKA, Anna  
 ; TITLE OF INVENTION: ESTERASES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FISH & RICHARDSON P.C.  
 ; STREET: 4225 EXECUTIVE SQUARE, STE 1400  
 ; CITY: LA JOLLA  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 INCH DISKETTE  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WORD PERFECT 6.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/027.804  
 ; FILING DATE: 21-Dec-2001  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/602,359  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: HALDE, LISA A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 09010/010001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-678-5070  
 ; TELEFAX: 619-678-5099  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 249 AMINO ACIDS  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: PROTEIN  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
 US-10-027-804-39  
 Query Match 70.8%; Score 901.5; DB 9; Length 249;  
 Best Local Similarity 71.5%; Pred. No. 4.3e-83;  
 Matches 176; Conservative 31; Mismatches 38; Indels 1; Gaps 1;  
 QY 1 LRLKPEINLVLSGGAAGTAHIGVLKAINELGIRVALSGVSAGAIVSFVYASGSP 60  
 DB 5 LRLKPEINLVLSGGAAGTAHIGVLKAINELGIRVALSGVSAGAIVSFVYASGSP 64  
 QY 61 GMSLLKRVNWLKFKPKPLKGLIGWEKAIKFLFVLPYRRIEKLPIPTIYCATDLYSG 120  
 DB 65 EMLKLEVNWLKFKPKPKMGLMGWEKAAEFLEKELGVKRLDNLNIPYLCSDLYTG 124  
 QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNNLVLDGGIVNNLPVPEPQSGIPTVCVDV 180  
 DB 125 KALYFGGDLIPVLLGSKSIFGIFEPVEYENFLVDGGINNLPVPEPKEFPIIGVDV 184  
 QY 181 LPIEPKDNKILHILRSFFFLAVRSNKKKFCDLVIVPELEBTTPLDVRKAQIWER 240  
 DB 185 LPTQBRKKINLHILRSFFFLAVRSNKKKFCNVIIPPLEEFSPLDVKNADEIF-C 243  
 QY 241 GYKAL 246  
 | : : :

DB 244 GDMRAL 249  
 RESULT 7  
 US-10-147-026-10  
 ; Sequence 10, Application US/10147026  
 ; Publication No. US20030003538A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dietrich, Paul Shartzer  
 ; APPLICANT: Huang, Chiao-Chain  
 ; APPLICANT: Johnson, Carl D.  
 ; APPLICANT: Sangameswaran, Lakshim  
 ; TITLE OF INVENTION: Neuropathic Pain Genes; Compositions  
 ; FILE REFERENCE: ROCH-006  
 ; CURRENT APPLICATION NUMBER: US/10/147,026  
 ; CURRENT FILING DATE: 2002-05-15  
 ; PRIOR APPLICATION NUMBER: 60/155,702  
 ; PRIOR FILING DATE: 1999-09-23  
 ; PRIOR APPLICATION NUMBER: 60/189,931  
 ; PRIOR FILING DATE: 2000-04-04  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 10  
 ; LENGTH: 425  
 ; TYPE: PRT  
 ; ORGANISM: rattus sp.  
 US-10-147-026-10  
 Query Match 11.0%; Score 139.5; DB 9; Length 425;  
 Best Local Similarity 23.2%; Pred. No. 1e-05;  
 Matches 64; Conservative 41; Mismatches 104; Indels 67; Gaps 12;  
 QY 10 NLVLSGGAAGTAHIGVLKAINELGIRV----RALSGVSAGAIVSFVYASG----YSPEG 61  
 DB 9 SLSFAGCGFLGFYHIGATLCLSERAPHILREARFFGCSAGALHAVTFVCSLPDLHIMEI 68  
 QY 62 MFSLLKRVNWLKFKPKPLKGLIGWEKAIKFLFVLPYRRIEKLPIPTIYCATDLYSG 120  
 DB 69 LMDLVKARSNRNIGTLHP-----FFNINKVCRDQGLQETLPDNVHIOIISGVYISLTVRS 134  
 QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNNLVLDGGIVNNLP-----164  
 DB 125 ENLVSEFHSKDEVDALVCSFCIFLPSGLIPP-SFRCERYVDGVDSDNVPLDAKTTIT 183  
 QY 165 VEPFQESGIPTVCVDVLPLE-PEKDIKNI-----LHILRSFFFLAVRSNKKKFC 215  
 DB 184 VSPFY--GERDICKVKSTNFIQVNTNLSLRLCTGNLHLLTRALF-----227  
 QY 216 DLVIVPELEBTTPLDVRKAQIWERGIYKALEVLSR 251  
 DB 228 -----PSDVKVMGELCFQGYLDAFRLEE 251  
 RESULT 8  
 US-09-765-205-40  
 ; Sequence 40, Application US/09765205  
 ; Patent No. US20020034800A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Li  
 ; TITLE OF INVENTION: BONE MARROW SECRETED PROTEINS AND POLYNUCLEOTIDES  
 ; FILE REFERENCE: 1458.004/200130.449  
 ; CURRENT APPLICATION NUMBER: US/09/765,205  
 ; CURRENT FILING DATE: 2001-01-17  
 ; PRIOR APPLICATION NUMBER: US/09/212,440  
 ; PRIOR FILING DATE: 1998-12-16  
 ; NUMBER OF SEQ ID NOS: 46  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 40  
 ; LENGTH: 504  
 ; TYPE: PRT  
 ; ORGANISM: human  
 US-09-765-205-40

## Query Match

Best Local Similarity 8.6%; Score 109.5; DB 10; Length 504;  
 Best Local Similarity 26.0%; Pred. No. 0.014;  
 Matches 69; Conservative 32; Mismatches 115; Indels 49; Gaps 12;

10 NIVLSGAAKGIAMHGVTKAINE-----IGIRVAAISGVSAGIVSVFYASGYSPEGMFSL 65  
 9 NISPGCGFLGYIVGVASCSREHAPLVANAHITIGASGALTRALVTVGCGAGAK 68  
 66 LKRVNMLKFKRPPLKGLIGMEKALR-FLIEVLPRRIEKLIPFYICADLYSGRLY 124  
 69 FLEVSKEARKFPLGLPHSPFLVLTISFLIKVLPADSHHAGSGRLGISTFVSGENVY 128  
 125 L-----SEGLIPALISCAIP---GIEP-----VEYKNYLVDSGIVNNLP----- 164  
 129 IGHFMSKDELIONAVCSGFIPIVYCGILPESLQGRY-----VDGISTDLPIYELKNTIT 183  
 165 VEPFQESGIPYVCVDVLEFIEPKDKITLHILRSFPLAVRSNREKREKFCDLVTPPLE 224  
 184 VSPF--SGESDIC-----FQDSSTIHR--LRVNTSIQFNLRLYRLSKALFPPE-- 230  
 225 EFTPLDVRKADQIMERGYIKALEVL 249  
 231 ---PLVLR---EMCKQGYRDLGLRFL 249

## RESULT 9

US-10-034-937-14  
 ; Sequence 14, Application US/10034937  
 ; Publication No. US20030097684A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Carr, Brian  
 ; APPLICANT: Rosen, Barbara A.  
 ; APPLICANT: Bermudez, Ericka R.  
 ; APPLICANT: Ness, Jon E.  
 ; APPLICANT: Maxygen, Inc.  
 ; TITLE OF INVENTION: Lipid Acyl Hydrolases and Variants Thereof  
 ; FILE REFERENCE: 18097A-032010US  
 ; CURRENT APPLICATION NUMBER: US/10/034,937  
 ; PRIOR FILING DATE: 2002-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/260,477  
 ; NUMBER OF SEQ ID NOS: 47  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 14  
 ; LENGTH: 390  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: clone p1P-10  
 ; OTHER INFORMATION: Improved pentin lipid acyl hydrolase  
 US-10-034-937-14

## Query Match

Best Local Similarity 6.8%; Score 86.5; DB 9; Length 390;  
 Best Local Similarity 22.7%; Pred. No. 2.1;  
 Matches 53; Conservative 30; Mismatches 79; Indels 71; Gaps 11;

11 LVLSGAAKGIAMHGVTKAINE-----ELGIRVAAISGVSAGIVSV----- 51  
 21 LAIDGGGIRGIIIPGILKQLEATLQRMDSARLAAYFDVAVAGSTGIIITALLTPADPQN 80  
 52 ---FYASG---YSPGMSFLKRVNMLKFRKRPPLKGLIGMEKALRFLIEVLPRRI 102  
 81 KDRPLVAGEIIDFIEHGPSEIFNKSSTACSSPGIFGCKYDG-----KYIQETIISQKL 132  
 103 IEKL-----ELPFIYICADLYSGRALYISGSL--IPAL-----LQSCAIPGIF 144  
 133 NETLDDQTTNVIPIPSF---DILKLRPTIFSFNKEVPELAVNKLSDVCMGTSAAPIVF 188  
 145 EEVEYKN-----YLVDGGIVNNLVEPFGESGIPYVCVDVLEFIEPKDKITL 193  
 189 PPIYRAGDTPEFLVDGAILADIFA-----FVALSEVLQOEYKN--KEIL 232

## RESULT 10

US-10-160-758-11  
 ; Sequence 11, Application US/10160758  
 ; Publication No. US20030036076A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EXELIXIS, INC.  
 ; TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
 ; FILE REFERENCE: EX02-083C  
 ; CURRENT APPLICATION NUMBER: US/10/160,758  
 ; CURRENT FILING DATE: 2002-06-03  
 ; PRIOR APPLICATION NUMBER: US 60/296,076  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/328,605  
 ; PRIOR FILING DATE: 2001-10-10  
 ; PRIOR APPLICATION NUMBER: US 60/357,253  
 ; PRIOR FILING DATE: 2002-02-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 11  
 ; LENGTH: 3354  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-160-758-11

## Query Match

Best Local Similarity 6.7%; Score 85.5; DB 9; Length 3354;  
 Best Local Similarity 21.3%; Pred. No. 54;  
 Matches 53; Conservative 36; Mismatches 75; Indels 85; Gaps 11;

22 AHIGVLR-----AINEIGIRVRA-----LSGVSAGIVSVFYASGYSPEGMFSL 65  
 565 AYVALRENEPSVTQL-VLIRATDESDPPNNQITTSIVSASAFSGSYDISGYGVISV 623  
 66 LKRVNMLKFKRPPLKGLI-----GMEKALRFLIEVLPRRI 99  
 624 SRPLDYEQISNGILYFTVAMDMGNPLNSTVPTIEVDENDNPFTSKRPAFVSIVV-- 681  
 100 YRIEKLIEPFIYICADLYSGR-----ALYISEGSL--IPALISCAIPGIFEPVEYK 150  
 682 -ENIMAGATVLPINATDLDRESEYQESIIYLSBSTOFRINASGEITTSLLDRETKS 740  
 151 NYLL-----VDGIVNNLVEPFGESGIPYVCVDVLEFIEPKDKITL 186  
 741 EYLIYRAVDGSGVGN-----OKTGLATVNTITLIDINDNHPYKADPYIINLVEMTPPD 794  
 187 KQINKILHI 195  
 795 SDVTTVVAV 803

## RESULT 11

US-10-160-758-12  
 ; Sequence 12, Application US/10160758  
 ; Publication No. US20030036076A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EXELIXIS, INC.  
 ; TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
 ; FILE REFERENCE: EX02-083C  
 ; CURRENT APPLICATION NUMBER: US/10/160,758  
 ; CURRENT FILING DATE: 2002-06-03  
 ; PRIOR APPLICATION NUMBER: US 60/296,076  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/328,605  
 ; PRIOR FILING DATE: 2001-10-10  
 ; PRIOR APPLICATION NUMBER: US 60/357,253  
 ; PRIOR FILING DATE: 2002-02-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 12  
 ; LENGTH: 3354  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-160-758-12



PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5379  
LENGTH: 391  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5379

Query Match 6.5%; Score 83; DB 10; Length 391;  
Best Local Similarity 23.2%; Pred. No. 4.7;  
Matches 62; Conservative 36; Mismatches 93; Indels 76; Gaps 14;

QY 32 ELGIRVALSGVS-AGAIVSVFYASGSPGMSLSLRVNWMLKLFKPKPLKGLIGMEKA 90  
DB 142 EAAVDLAKLTGAKPAGVICEIMNDGTWAKGD-----LQFKERKQLKMIT----- 188  
QY 91 IRPLEEVLPRR-----IE---KLEIPYICATDLYSGRALYLSBSLI----- 131  
DB 189 ---IDDLIEYRKKEPEIEFKAKVMPTDFGTFMVGFKATYTDERIVLTKGAIROHEN 245  
QY 132 PALSGCAIRGIF-----EPVEYKN-----YLVVG---GIVNNLPVPEPQ 169  
DB 246 VRLHSACLTGDI FHSQRCDGAGQLESMSKYNIEHGMIITLPQEGIGIGLNLKRAYELI 305  
QY 170 ESGIPYVCVDVLPRIPEKDIK-----NIIHLIRSFPLAVSNSSEKKEKCDL 217  
DB 306 EGGYDVTYTN-LALGPDDELDRDHYHIAQLKTFNIEHINLS---NNPSKFEGLKQY-GI 360  
QY 218 VYVPELSEFTPLDYKADQIMERGYIK 244  
DB 361 DIARIEVIVPEYVHNHD-YMETKKIK 386

## RESULT 15

US-09-815-242-12285  
Sequence 12285, Application US/09815242  
Patent No. US2002006159A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12285  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12285

Query Match 6.5%; Score 83; DB 10; Length 393;  
Best Local Similarity 23.2%; Pred. No. 4.7;  
Matches 62; Conservative 36; Mismatches 93; Indels 76; Gaps 14;

QY 32 ELGIRVALSGVS-AGAIVSVFYASGSPGMSLSLRVNWMLKLFKPKPLKGLIGMEKA 90  
DB 144 EAAVDLAKLTGAKPAGVICEIMNDGTWAKGD-----LQFKERKQLKMIT----- 190  
QY 91 IRPLEEVLPRR-----IE---KLEIPYICATDLYSGRALYLSBSLI----- 131  
DB 191 ---IDDLIEYRKKEPEIEFKAKVMPTDFGTFMVGFKATYTDERIVLTKGAIROHEN 247  
QY 132 PALSGCAIRGIF-----EPVEYKN-----YLVVG---GIVNNLPVPEPQ 169  
DB 248 VRLHSACLTGDI FHSQRCDGAGQLESMSKYNIEHGMIITLPQEGIGIGLNLKRAYELI 307  
QY 170 ESGIPYVCVDVLPRIPEKDIK-----NIIHLIRSFPLAVSNSSEKKEKCDL 217  
DB 308 EGGYDVTYTN-LALGPDDELDRDHYHIAQLKTFNIEHINLS---NNPSKFEGLKQY-GI 362  
QY 218 VYVPELSEFTPLDYKADQIMERGYIK 244  
DB 363 DIARIEVIVPEYVHNHD-YMETKKIK 388

Search completed: June 19, 2003, 15:13:00  
Job time: 183 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 14:22:25 ; Search time 21 Seconds  
(without alignments)  
495.741 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273  
Sequence: 1 LRLRRFEINILVSGGAAG.....RKADQIMRGYKALAEVLSR 251

Scoring table: BLOSUM62  
Gapco 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt 40.1\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	268.5	21.1	260	YLBK_BACSU	Q34731 bacillus su
2	257.5	20.2	314	YCHK_BCOLI	P37053 escherichia
3	242	19.0	1679	YMF9_YEAST	Q04958 saccharomyc
4	237.5	18.7	192	YCHK_SHIFL	P37054 shigella fl
5	235	18.5	583	YPE5_MYCTU	O50733 mycobacteri
6	234	18.4	360	YAE3_MYCTU	O54411 mycobacteri
7	223.5	17.6	291	YOH0_BACSU	P54513 bacillus su
8	217.5	17.1	1425	SMS_DROME	O94969 drosophila
9	208.5	16.4	1371	YOL4_CAEEL	Q02331 caenorhabdi
10	189.5	14.9	880	YU7_CAEEL	Q21534 caenorhabdi
11	132.5	10.4	910	YK69_YEAST	P31653 saccharomyc
12	124	9.7	621	YPO7_COEEL	Q11186 caenorhabdi
13	121	9.5	253	G82_HUMAN	P41247 homo sapien
14	100.5	7.9	357	YJUT_ZOCCA	P39407 escherichia
15	98.5	7.7	435	FIXC_AZOCA	P26484 azorhizobiu
16	98.5	7.7	1073	YCP4_YEAST	P43564 saccharomyc
17	94.5	7.4	769	HCPA_YEAST	O31819 thauera arc
18	92	7.2	531	FNX1_SCHRO	Q03752 schizosach
19	92	7.2	786	GE2P_HUMAN	P41002 homo sapien
20	92	7.2	1102	YE20_METUA	Q58813 methanococc
21	91	7.1	759	PURL_THEVO	Q97bd5 thermoplas
22	90	7.1	469	SYEL1_THEMA	O94172 thermotoga
23	89.5	7.0	588	CALI_HUMAN	Q11939 homo sapien
24	87.5	6.9	397	HEM1_PYPAR	Q84YV6 pyrobaculum
25	86.5	6.8	282	DH11_MOUSE	P05177 mus musculu
26	86.5	6.8	4427	PKS1_BACSU	Q05470 bacillus su
27	86	6.8	456	PROA_YEAST	P54885 saccharomyc
28	86	6.8	1161	DP3A_AQUAR	O67125 aquifex aeo
29	85.5	6.7	3354	CADN_HUMAN	Q91251 homo sapien
30	84	6.6	255	UPPS_METTH	O26334 methanobact
31	83.5	6.6	431	SYD_METKA	O8XG34 methanopyru
32	83.5	6.6	564	TP6B_PYPAR	O9Y135 pyrococcus
33	83.5	6.6	773	SYFB_CAMDE	O9Pp35 campylobact

34	83	6.5	1884	RRP5_HUMAN	Q14690 homo sapien
35	82.5	6.5	292	DH11_HUMAN	P28845 homo sapien
36	82.5	6.5	435	FIXC_BRAJA	P10331 bradyrhizob
37	82.5	6.5	564	TP6B_PYPAR	O74020 pyrococcus
38	82.5	6.5	698	ALT_BPT2	Q38424 bacterioph
39	82	6.4	242	TRUA_HBPLU	Q92K00 helicobacte
40	82	6.4	496	SRM_MOUSE	Q62270 mus musculu
41	82	6.4	948	SECA_SYNP7	O55357 synchococc
42	81.5	6.4	318	TEG_BACME	P30420 bacillus me
43	81.5	6.4	319	KGF_THEMA	O9WY32 thermotoga
44	81	6.4	729	SYG_MOUSE	Q9CZD3 mus musculu
45	81	6.4	759	NAH2_SCHPO	Q14123 schizosach

## ALIGNMENTS

RESULT 1  
ID YLBK\_BACSU STANDARD; PRT; 260 AA.  
AC 034731;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein ylbk.  
GN YLBK.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Berrero M., Prescan E., Glaser P., Richou A., Danchin A.;  
RT "Bacillus subtilis chromosomal region downstream mpbE";  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azeredo V., Berrero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Bourelier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Comerion I.F., Cummings N.J., Daniel R.A.,  
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Eutelan K.D., Ewington J., Fabret C., Ferrati E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Goltschky E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Hentut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,  
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,  
RA Prescan E., Fujic P., Fumelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scotlone F.,  
RA Sekiguchi J., Sekowka A., Setor S.J., Setor P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takauchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
RA Toasato V., Uchiyama S., Vandenbol M., Vanter F., Vassarotti A.,  
RA Viari A., Wandut R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters E., Wipat A., Yamamoto H., Yamane K., Yamamoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis".  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SMS) FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



DB 238 AVTAPATETITSTISIVLEKRLKRNMGADPPDIILQVCCPISITLDPHNAIANGOL 237  
 QY 244 KALEVISE 251  
 DB 298 AVERKME 305

RESULT 3  
 YMF9\_YEAST STANDARD; PRT; 1679 AA.

AC Q04958;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 187.1 kDa protein in OG4-CNA2 intergenic region.  
 GN YML059C OR YMW958.03C  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Baktariota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288c / AB972;  
 RA Devlin K., Churcher C., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SMS) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL; Z46729; CA86716.1; -  
 CC SGD; S0004524; YML059C.  
 DR InterPro; IPR002641; Patatin.  
 DR InterPro; IPR001423; UPF0028.  
 DR InterPro; IPR000595; CNMP binding.  
 DR Pfam; PF00027; CNMP binding; 2.  
 DR SMART; SM01734; Patatin; 1.  
 DR SMART; SM00100; CNMP; 1.  
 DR PROSITE; PS00042; CNMP BINDING\_3; 2.  
 DR PROSITE; PS01237; UPF0028; 1.  
 DR Hypothetical protein; Transmembrane.  
 KM TRANSMEM 50 70 POTENTIAL.  
 FT TRANSMEM 104 124 POTENTIAL.  
 SO SEQUENCE 1679 AA; 187132 MW; 550PFCDAACAF8E25 CRC64;

Query Match 19.0%; Score 242; DB 1; Length 1679;  
 Best Local Similarity 28.5%; Pred. No. 1.2e-13;  
 Matches 79; Conservative 42; Mismatches 120; Indels 36; Gaps 6;

QY 1 LRLKRF---EINIVLSGGAAGIAHIGVLAINEIGIRVRLSGVSAGATVSFVAGSY 57  
 DB 1360 LRLARILSGAIGVAGGARGISHLGIVQINIEGIPVDVIGTSIGSVGLVAKDY 1419

QY 58 SPBGMFSLK---RVN-WLKIFKRPPLKGLIGKEKAIPLIEVLPYRIEKLPTPT 111  
 DB 1420 DLVPIYGRVKKFAGRISSIRMLTDLTWPTVSTYTGHEFRNRGIWTFPGDTIEIDPFWIOY 1479

QY 112 ICATDLYSGALYLSGSLIPALGSCAIPGIFEPVEYKNTLLVDGIVNNLPVEPFOES 171  
 DB 1480 CNGSNINIDSVQEHISFQYARIRYIRAWMSLAGLPLLENGSWLDDGGVVDLPTFEMBAR 1539

QY 172 GIPTV-CVDV-----LPIPEKDIKNILHILLSFPFLAVR 205  
 DB 1540 GCQTIPADVGSADRTWMEYGSLSNGFWIIFRRANPFSSHPNIDMAEIQVRLGYAVSV 1599

QY 206 SNSEKREKCDLVIV-PELEEPTPLDVRKADIMERG 241

DB 1600 MALEKATNGVYVVRPIEBYATIDFSKFEIYHNG 1636

RESULT 4  
 YCHK\_SHIFL STANDARD; PRT; 192 AA.

AC P37054;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 34.4 kDa protein in HNR-PUR0 intergenic region  
 DE (Fragment).  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Serotype 2A;  
 RX MEDLINE=93023838; PubMed=1406252;  
 RA Hromockyj A.E., Tucker S.C., Maurelli A.T.;  
 RT "Temperature regulation of Shigella virulence: identification of the  
 RT repressor gene vtr, an analogue of hns, and partial complementation  
 RT by tyrosyl transfer RNA (tRNA(Tyr)).";  
 RL Mol. Microbiol. 6:2113-2124(1992).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Rudd K.E.;  
 RL Unpublished observations (FEB-1994).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SMS) FAMILY.  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO  
 BE INTRODUCED IN POSITION 178 TO PRODUCE THIS ORF.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL; X66849; -; NOT\_ANNOTATED\_CDS.  
 DR InterPro; IPR002641; Patatin.  
 DR InterPro; IPR001423; UPF0028.  
 DR Pfam; PF01734; Patatin; 1.  
 DR PROSITE; PS01237; UPF0028; 1.  
 KM Hypothetical protein.  
 FT NON TER 192  
 SO SEQUENCE 192 AA; 20831 MW; 34AEFGA693CC917A CRC64;

Query Match 18.7%; Score 237.5; DB 1; Length 192;  
 Best Local Similarity 33.7%; Pred. No. 2.1e-14;  
 Matches 58; Conservative 33; Mismatches 64; Indels 17; Gaps 4;

QY 3 LRKEEINVLGSGAAGIAHIGVLAINEIGIRVRLSGVSAGATVSFVAGSYPEGM 62  
 DB 14 MKRI-KIALALGSGAARGSHIGVINALKKVETIEDIYAGSISIVGAAYAC-----DR 67

QY 63 FSLIKRVNLIKLFKFPYK-----GLIGWEKAIPLIEVLPYRIEKLPTPT 113  
 DB 68 LSALE-DWTSFYSYELLRLMDLWSORGLRGRRVNOYREIMPETEIENGSRPAPV 125

QY 114 ATDLYSGALYLSGSLIPALGSCAIPGIFEPVEYKNTLLVDGIVNNLPV 165  
 DB 126 ATNLSIRELWTFEGDLHLAIRASCSIRGLMAPVAHNYMTWVDGAVNPIPI 177

RESULT 5  
 YP65\_MYCTU STANDARD; PRT; 583 AA.  
 AC Q50733;  
 DT 01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)  
 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV2565  
 GN RV2565 OR MT2641 OR MT1934.03C.  
 OS Mycobacterium tuberculosis  
 OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultston J.E., Taylor K., Whitthead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; Z77250; CAN01050.1; -  
 CC EMBL; AE007098; AAK46954.1; -  
 CC TIGR; MT2641; -  
 CC TubercuList; RV2565; -  
 CC InterPro; IPR002641; Patatin.  
 CC InterPro; IPR001423; UPF0028.  
 CC Pfam; PF00027; cNMP binding.  
 CC Pfam; PF01734; cNMP binding; 1.  
 CC SMART; SM00100; cNMP; 1.  
 CC PROSITE; PS0042; cNMP BINDING 3; 1.  
 CC PROSITE; PS01237; UPF0028; 1.  
 CC Hypothetical protein; Complete proteome.  
 FT CONFLICT 375 375 G -> D (IN REF. 2).  
 SQ SEQUENCE 583 AA; 62123 MW; EC6EE062BAEC62F7 CRC64;  
 Query Match 18.5%; Score 235; DB 1; Length 583;  
 Best Local Similarity 24.7%; Pred. No. 1.4e-13;  
 Matches 70; Conservative 51; Mismatches 106; Indels 56; Gaps 6;  
 QY 9 INLVGGGAAGIAHIGVKAINEIGIRVRLSGVAGAVSVFYA----- 54  
 DB 307 LGLVMAGGARGAIAHFGVYQELTAGVVIDRFGTSSGAISAAAFALMGADGADIAAARE 366  
 QY 55 --SGVSPEGMFLLKRVNVLKFKPKPLKGLIGWEKATRFLEVLPRYRIKLEIPIYI 112  
 DB 367 FIAGSDPLADYII-----PISALTRGGRVRLVGVFGVNTLIEHLPRGFFS 412  
 QY 113 CATDLYSGRAVLYSGSLIPALGSCAIPGIFPEVEYKNVLLVGGIVNNLPVPEPF-QES 171

DB 413 VSADMITGQDIHRRGVSAGVRAISIFGLPPVHNGEQLLVGGLLNNLPANVMCACT 472  
 QY 172 GIFTVCYD-----VLP--IEPEKIKNLL-----HILRSFFLAURS 206  
 DB 473 DGEVICVDLARTFTVPSKGLGPPVTPGGLLRLLLTCTDNALPLOTETLLRADFLAANT 532  
 QY 207 NSERKKEFCDLVTPLEBFTPLDVRKADQIMERGVIKALEVL 249  
 DB 533 ANLRELPRVAATIEPDVSKIGLVNFKQIDAALAEAGRMAAAL 575  
 RESULT 6  
 YA63 MYCTU  
 ID YAG3 MYCTU STANDARD; PRT; 360 AA.  
 AC O53411;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV1063c.  
 GN RV1063C OR MT1093 OR MT017.16C.  
 OS Mycobacterium tuberculosis  
 OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultston J.E., Taylor K., Whitthead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AL021897; CAN17179.1; -  
 CC EMBL; AE006990; AAK45347.1; -  
 CC TIGR; MT1093; -  
 CC TubercuList; RV1063c; -  
 CC InterPro; IPR002641; Patatin.  
 CC InterPro; IPR001423; UPF0028.  
 CC Pfam; PF01734; Patatin; 1.  
 CC PROSITE; PS01237; UPF0028; FALSE NEG.  
 CC Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 360 AA; 37522 MW; F08C30CE6096D0F6 CRC64;  
 Query Match 18.4%; Score 234; DB 1; Length 360;





Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champagne M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Arbil J.F., Agayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 Posler K., Gabriellian A.E., Garg N.S., Gelbart M.M., Gleaser K.,  
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Waesman D.A., Weinstein G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE SIGNALING MECHANISM BETWEEN  
 CC NEURONS AND GLIA THAT REGULATES GLIA WRAPPING DURING DEVELOPMENT  
 CC OF THE ADULT BRAIN.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: BOTH ISOFORMS ARE EXPRESSED IN CORTICAL CELL  
 CC BODIES OF ADULT BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: THE LONG ISOFORM IS EXPRESSED IN ALL  
 CC DEVELOPMENTAL STAGES WITH HIGHEST LEVELS IN YOUNG EMBRYOS AND  
 CC ADULTS. THE SHORT ISOFORM IS DETECTED ONLY IN ADULT  
 CC HEAD.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.  
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z97187; CA951772.1;  
 CC ENBL; A5003442; AAF46305.1; ALT\_SEQ.  
 CC Flybase; FBgn0003656; sws.  
 CC InterPro; IPR002641; Patatin.  
 CC InterPro; IPR001423; UPF0028.  
 CC InterPro; IPR000595; CNMP binding.  
 CC Pfam; PF00027; CNMP binding; 3.  
 CC Pfam; PF01734; Patatin; 1.  
 CC SMART; SM00100; CNMP; 2.  
 CC PROSITE; PS50042; CNMP\_BINDING\_3; 3.  
 CC PROSITE; PS01237; UPF0028; 1.  
 CC Alternative 648 648  
 CC G->R: IN ALLELE SWS-5; AGE-DEPENDENT  
 CC MUTAGEN 648 648  
 CC G->D: IN ALLELE SWS-4; AGE-DEPENDENT  
 CC MUTAGEN 956 956  
 CC G->D: IN ALLELE SWS-4; AGE-DEPENDENT  
 CC FT

FT SQ SEQUENCE 1425 AA; 160594 MW; 2A7AF2DAC06FE53 CRC64;  
 NEURODEGENERATION.  
 Query Match 17.1%; Score 217.5; DB 1; Length 1425;  
 Best Local Similarity 26.1%; Pred. No. 1.6e-11;  
 Matches 73; Conservative 44; Mismatches 108; Indels 55; Gaps 9;  
 QY 9 INLVSGGAAGIAHIGVLRKALNELGIRVALSGVSAIGAVSVFYASGYSPEGMFSL-- 66  
 DB 950 IGLVLGGGARGAAHIGMLKAIQAGIPVDMVGVSIGALMGALWCS---ERNITVTQ 1005  
 QY 67 -----KRVNW-LKLKFKPPKGLGWEKARFLEELVPRRIEKLPTVICADLY 118  
 DB 1006 KARWSKQWYKWLQDLDTPTYSMFSGREFNKTHTDFTGDSISDLMWIPYTLTTDT 1065  
 QY 119 SGRALYISGLIPALGLSCAIPGIFPV--EYKNVLLVDGIVNNLPVEPFOESGIP-T 175  
 DB 1066 ASCHRIHTNGSLWYVRSSNSLSGYMPLCDPKDGHLLDGGVYNNLPADVHNLGAHI 1125  
 QY 176 VCVDV-----LPIEPKDKNHLILHLSRFLAV 204  
 DB 1126 IADVGSODDTLDNYGDDLSGWMLLYKNWPTSPVKVPDLPDIQSLR-----AYSCV 1180  
 QY 205 RSNSE-KRKEPCDLIVPELEEFPLDVRKADQIMERGI 243  
 DB 1181 RQLEEVKNSDYCE-YIRPPIDKYKTLAFSGSFDEIRDVG 1219  
 RESULT 9  
 YOLA CAEL  
 ID YOLA CAEL STANDARD; PRT; 1371 AA.  
 AC Q02331; Q95PX1;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ZK370.4 in chromosome III.  
 GN ZK370.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones A., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sonnenhammer E., Staden R.,  
 RA Sultat J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RA Waterston R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a and b (shown here); may be  
 CC produced by alternative splicing.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL; M98552; AAA28206.2; -  
 DR EMBL; M98552; AAL11481.1; -  
 DR WormPep; ZK370.4a; CE25687.  
 DR WormPep; ZK370.4b; CE29641.  
 DR InterPro; IPR002641; Patatin.  
 DR InterPro; IPR001423; UPR0028.  
 DR InterPro; IPR000595; cNMP binding.  
 DR Pfam; PF00027; cNMP binding; 3.  
 DR Pfam; PF01734; Patatin; 1.  
 DR SMART; SM00100; cNMP; 2.  
 DR PROSITE; PS50042; cNMP BINDING\_3; 3.  
 DR PROSITE; PS01237; UPR0028; 1.  
 KM Hypothetical protein: Alternative splicing.  
 FT VARSPLIC 512 522 MISSING (IN ISOFORM A).  
 SQ SEQUENCE 1371 AA; 154528 MW; DADP21707BD47782 CRC64;  
 Query Match 16.4%; Score 208.5; DB 1; Length 1371;  
 Best Local Similarity 27.1%; Pred. No. 1e-10;  
 Matches 75; Conservative 39; Mismatches 110; Indels 53; Gaps 9;  
 QY 9 INVLGGAAGKAIHIGVKAINEIGIRVRLSGVSGAIVSVFYASGSP----- 60  
 DB 958 IGVLVGGGAKGAAHAGVLRALREKGVYDVGSTISGLGUYAR--TPDIVVYETRA 1015  
 QY 61 -----GMSFLKRVNMLKFKRPPKGLIGMEKAIKFLBEVLPYRRIEKLIPYICAT 115  
 DB 1016 ASFWNGMSL-----WRKLLDTVAHSAHFAGQNFISIKDLFEERLIEDLWISYFCIST 1070  
 QY 116 DLVSGRALYVSESSLIIPALLGSCAIPGIFEPV--EYKNYLLVDGIVNNLPVPEPQSG 173  
 DB 1071 DISTSEKVRKRSPLVAYGRASMSLAGTIPICDPDGHLLDGGYVNNVPADVMENLQA 1130  
 QY 174 P-TVCDVLPPIR-----PEKDKNLIHILSFFLAVS 206  
 DB 1131 RCVIACDVGSIEETNLVDYDGLSGMWVLLKRLNFGTTPPR--IIMMEIOSRLIAYSCVR 1189  
 QY 207 NSR--KRKECDLVIVPELBEFTPLDVKRAQDIEMRG 241  
 DB 1190 QLEVVKKASYC-RYLKRPPIEPKTIIDFKFQGIHMLG 1225  
 RESULT 10  
 ID YVL7\_CABEL STANDARD; PRT; 880 AA.  
 AC Q21534; Q20023;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein M10.7 in chromosome II.  
 OS M10.7.  
 GN Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thomas K., Baynes C.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPR0028 (SMS) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; Z49966; CAA90246.1; -  
 DR EMBL; Z49968; CAA90246.1; JOINED.

DR EMBL; Z49968; CAA90246.1; -  
 DR EMBL; Z49966; CAA90246.1; JOINED.  
 DR WormPep; M10.7; CE03510.  
 DR InterPro; IPR002641; Patatin.  
 DR InterPro; IPR001423; UPR0028.  
 DR InterPro; IPR000595; cNMP binding.  
 DR Pfam; PF00027; cNMP binding; 1.  
 DR Pfam; PF01734; Patatin; 1.  
 DR SMART; SM00100; cNMP; 2.  
 DR PROSITE; PS50042; cNMP BINDING\_3; 1.  
 DR PROSITE; PS01237; UPR0028; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 880 AA; 100053 MW; A94D9EFED3A7DEED CRC64;  
 Query Match 14.9%; Score 189.5; DB 1; Length 880;  
 Best Local Similarity 23.0%; Pred. No. 3e-09;  
 Matches 65; Conservative 45; Mismatches 113; Indels 59; Gaps 8;  
 QY 9 INVLGGAAGKAIHIGVKAINEIGIRVRLSGVSGAIVSVFYASGSP----- 60  
 DB 551 IGVLVGGGAKGAAHAGVLRALREKGVYDVGSTISGLGUYAR--TPDIRAVGRMK 608  
 QY 61 -----GMSFLKRVNMLKFKRPPKGLIGMEKAIKFLBEVLPYRRIEKLIPYICAT 111  
 DB 609 DFTDRLRNLIIDVRLDTM-----PYGILTGKRPVLCYQRLNDVNIEDCVSFF 660  
 QY 112 ICATDLYSGRALYVSESSLIIPALLGSCAIPGIFEPV--EYKNYLLVDGIVNNLPVPEPQ 169  
 DB 661 CITTDLTSSSMKIHNGIMVPPVSSMSIAGYVPICDPDGHLLDGGYVNNLPADIMR 720  
 QY 170 ESGIPV-CVDV-----LPIEPEKDKNLIHILSFFLA 202  
 DB 721 SLGANNVYIVDGMDDNTNKAAGFSSGTCLEKRWPFGEHIVLVNNVQKRLAYV 780  
 QY 203 AVRSNSE--KRKECDLVIVPELBEFTPLDVKRAQDIEMRG 242  
 DB 781 CCVOMEIVKNAQCYVYLP-IRSGIFDPSKPDQAQIGY 821  
 RESULT 11  
 ID YK69\_YEAST STANDARD; PRT; 910 AA.  
 AC P36165;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 102.7 kDa protein in PRP16-SR640 intergenic region.  
 GN YK0898C OR YK0409.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINB=94262327; PubMed=8203164;  
 RA Garcia-Cantalejo J., Baladron V., Estreban P.F., Santos M.A., Bou G.,  
 RA Remacha M.A., Revuelta J.L., Ballesca J.P.G., Jimenez A., del Rey F.;  
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces  
 RT cerevisiae chromosome XI contains the HBS1, MRP-120 and PRP16 genes,  
 RT and six new open reading frames".  
 RL Yeast 10:231-245 (1994).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: TO YEAST YMR313C AND S.POMBE SPC1450.16C.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; Z27116; CAA81640.1; -

```

DR EMBL; Z28314; CAA82168.1; --
DR FJR; S38167; S38167.
DR FJR; S39130; S39130.
DR SCD; S0001797; YR0089C.
DR InterPro; IPR002641; Patatin.
DR Pfam; PF01734; Patatin; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 282 302 POTENTIAL.
FT TRANSMEM 310 326 POTENTIAL.
FT TRANSMEM 426 442 POTENTIAL.
SQ SEQUENCE 910 AA; 102716 MW; 1CF03C4A6E64B9C CRC64;

Query Match
Best Local Similarity 10.4%; Score 132.5; DB 1; Length 910;
Best Local Similarity 30.4%; Pred. No. 0.00045;
Matches 62; Conservative 22; Mismatches 75; Indels 45; Gaps 9;

QY 11 LVLSGGAAGIAHIGVKAINELGIRVRALSGVSAIAIVSVFYASGYSPEGMFSLL--- 66
DB 282 LVLSGGTFLGFLHIGVGLTFLDLPVVISGSSAGAIVASI-LSVHHKEIPVLLNHIL 340

QY 67 -KRVNWLKLFKPKPLKGLI-----GW-----EKAIRFLEVL---PYRIEK 105
DB 341 DKFPIFKDDKQKSESNLTKISRFKNGTWFNDKHLVNTWIEFLGOLTFREAYNRGK 400

QY 106 LEIPYVICATDLYSGRAL--YLSGSLI-PALLGSCAIFGIF--RPVEYK----- 150
DB 401 ILNITVSPASLFEQPRLLNNLTAPNVLIWSAVCAVCSLPGIFPSSPLYEKOPKTCERKPW 460

QY 151 ----NVLVDGGIVNNLVEPFQF 170
DB 461 TGSSSVKFGVDSVNDLPISRLSE 484

RESULT 12
ID YPD7 CABEL STANDARD; PRT; 621 AA.
AC Q11136;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein C05D11.7 in chromosome III.
GN C05D11.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO HUMAN GS2 PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U00048; AA053832.2; --
CC Wormpep; C05D11.7; CE29663.
CC InterPro; IPR002641; Patatin.
CC Pfam; PF01734; Patatin; 1.
KW Hypothetical protein.
SQ SEQUENCE 621 AA; 69363 MW; 93BE9A868F684EC CRC64;

Query Match
Best Local Similarity 9.7%; Score 124; DB 1; Length 621;

DR EMBL; U03986; AA16491.1; --
DR EMBL; U08893; AA17838.1; --
DR EMBL; U08888; AA17838.1; JOINED.
DR EMBL; U08889; AA17838.1; JOINED.
DR EMBL; U08890; AA17838.1; JOINED.
DR EMBL; U08891; AA17838.1; JOINED.
DR EMBL; U08892; AA17838.1; JOINED.
DR MIM; 300102; --
DR InterPro; IPR002641; Patatin.
DR Pfam; PF01734; Patatin; 1.
SQ SEQUENCE 253 AA; 27980 MW; 6FFA94F460BA03AC CRC64;

Query Match
Best Local Similarity 9.5%; Score 121; DB 1; Length 253;
Best Local Similarity 23.9%; Pred. No. 0.001;
Matches 64; Conservative 47; Mismatches 109; Indels 48; Gaps 14;

QY 7 BEINLVSGGAAGIAHIGVKAINELGIR---VRALSGVSAIAIVSVFYASGYSPEGM 62
DB 2 KHINLSFAACGFLGIYHLGAASALCRHGKLVKDKVAPAGASAGSLVASVLLT--APE-- 57

Best Local Similarity 27.6%; Pred. No. 0.0016;
Matches 58; Conservative 28; Mismatches 72; Indels 52; Gaps 11;

QY 7 BEINLVSGGAAGIAHIGVKAINELG---IRVRALSGVSAIAIVSVFYASGYS-PEGMP 63
DB 9 ELMNLSFGCGFLCVYHAGVAAAIKEYAPELQNKILGASAGSIYACGLITGVCIASHATS 68
QY 64 SLKRVNWLKLFKPKP-----PLKGLIGWEKAIRFLEVLVPRRIEKLIEPTTYICATDLY 118
DB 69 TILKVSQARSRTFGPLHPEFNLGLVRDE-----LEVILP-----PN--AYEMC 111

QY 119 SGRAL-----YLSGSLIPALLGSCAIP-----GIFEPVEYKYNLLVDGGI 159
DB 112 TGLRVISLRTWSDHENVIIDEYRSNADLIDAIMCSFIPLYCGI-TTPKFRGVQYIDGGV 170

QY 160 VNNLVEPFQFQESGIPTVCDVLPFIEPKDI 189
DB 171 SDNQPI--YDEH-----TIVTSFFSGESDI 193

RESULT 13
GS2_HUMAN STANDARD; PRT; 253 AA.
ID GS2_HUMAN
AC P41247;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GS2 protein (DXS1283E).
GN GS2.
OS Homo sapiens (Human).
OC Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=95104848; PubMed=7806223;
RA Lee W.-C., Salido E., Yen P.H.;
RT "Isolation of a new gene GS2 (DXS1283E) from a CpG island between STS
RL Genomics 22:372-376(1994).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED, INCLUDING
CC HEART, BRAIN, PLACENTA, LUNG, LIVER, MUSCLE, KIDNEY, PANCREAS AND
CC SPLEEN.
CC -1- SIMILARITY: TO C.ELEGANS C05D11.7.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U03986; AA16491.1; --
CC EMBL; U08893; AA17838.1; --
CC EMBL; U08888; AA17838.1; JOINED.
CC EMBL; U08889; AA17838.1; JOINED.
CC EMBL; U08890; AA17838.1; JOINED.
CC EMBL; U08891; AA17838.1; JOINED.
CC EMBL; U08892; AA17838.1; JOINED.
CC MIM; 300102; --
CC InterPro; IPR002641; Patatin.
CC Pfam; PF01734; Patatin; 1.
SQ SEQUENCE 253 AA; 27980 MW; 6FFA94F460BA03AC CRC64;

Query Match
Best Local Similarity 9.5%; Score 121; DB 1; Length 253;
Best Local Similarity 23.9%; Pred. No. 0.001;
Matches 64; Conservative 47; Mismatches 109; Indels 48; Gaps 14;

QY 7 BEINLVSGGAAGIAHIGVKAINELGIR---VRALSGVSAIAIVSVFYASGYSPEGM 62
DB 2 KHINLSFAACGFLGIYHLGAASALCRHGKLVKDKVAPAGASAGSLVASVLLT--APE-- 57

```

QY 63 FSLIKKVNMLKLEKPEKPK-----GLI--GWEKAIKRP---LEEVYPRIRIEKLEPIFYIC 113  
 Db 58 --KIEKCNOR-TYKFAEIRIRSGFAGVATPGVDFMARLRSKESLIPSAHLEAQLRLVLS 114  
 QY 114 ATDLYSGR---ALYLSBSGLIPALIGSCAIP---GIFEPVEYKRYLLVDDGIIVNNLPVE 166  
 Db 115 ITAKATRENNHVLVTFSSRBDLIKVLASSFVPIYAGL-KLVEYKQKQKVDGGLNNALPIL 173  
 QY 167 PPOESGIPYCVVULPIBEKXI-----KNILHILRSFPLAVRSSEKREKFCOLVIVPE 222  
 Db 174 PVKRT-----VITSPFSGRLDISPODKQGLDLYV-----NIAKQDITMSLANDVR 218  
 QY 223 LEE-FTPLDVRKADQIMERGYIKALEVL 249  
 Db 219 LNOALFPPSKRKMEISLYCGFDDTVKFL 246

## RESULT 14

YJUU\_ECOLI STANDARD; PRT; 357 AA.  
 AC P39407;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yj31u.  
 GN YJUU OR B4337.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=95334362; PubMed=7610040;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U14003; AAA97273.1; -.  
 DR EMBL; AB000508; AAC77330.1; -.  
 DR Ecogene; BG12597; yj31u.  
 DR InterPro; IPR002641; Patatin.  
 DR Pfam; PF01734; Patatin; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 357 AA; 39831 MW; D7A57185B93CCEAA CRC64;

Query Match 7.9%; Score 100.5; DB 1; Length 357;  
 Best Local Similarity 26.5%; Pred. No. 0.11;

Matches 56; Conservative 30; Mismatches 72; Indels 53; Gaps 12;

QY 1 LRLRKEF-EINIVLSGGAAGIAHIGVAKINEIGIRVRA-----LSGVAGA-IVS 50  
 Db 15 LSLRPFQPGRIALVCEGGGGRGIFTAGVILDEF-----MRAQPNFDDYLTGTSAAQMTLS 68  
 QY 51 VEVAS--GSPPEGFSLIKRVNMLKLFKPKPLK-----GLIGWEKAIKRFLEEVLPYRRI 103  
 Db 69 AFICNGQGVARKYI-----MRYTTRREFPDRFRFGSGNLTIDLMVYEARSGMP----- 118  
 QY 104 EKEIIPF-----YCA--TDLYSGRALVYSEGLIPALIGSCAIPGIFE---FV 147  
 Db 119 --LQMDTARLFDGSKSFYWCACRODYPANFYFLPTKQNMVLIVIRASSAIPIGFYRSGVSL 176.

QY 148 EYKNVLLVDDGIIVNNLPVEPPOESGIPYCV 178  
 Db 177 EGINTL--DGGISDAIVKARAGAKTIVV 205

## RESULT 15

FIXC\_AZOCA STANDARD; PRT; 435 AA.  
 AC P26484;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FIXC protein.  
 GN FIXC.  
 OS Azorhizobium caulinodans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Hyphomicrobium group; Azorhizobium.  
 OX NCBI\_TaxID=7;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORS571;  
 RX MEDLINE=91203829; PubMed=1850088;  
 RA Arigoni F., Kaminski P.A., Hennecke H., Elmerich C.;  
 RT "Nucleotide sequence of the fixABC region of Azorhizobium caulinodans  
 RT ORS571: similarity of the fixB product with eukaryotic flavoproteins,  
 RT characterization of fixX, and identification of nifW";  
 RL Mol. Gen. Genet. 225:514-520(1991).  
 CC  
 CC -1- FUNCTION: COULD BE REQUIRED FOR THE FORMATION OF A FUNCTIONAL  
 CC NITROGENASE FE PROTEIN. PROBABLY ACCEPTS ELECTRONS FROM FIXA/FIXB  
 CC AND REDUCES A QUINONE.  
 CC  
 CC -1- COFACTOR: FAD (POTENTIAL).  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE ETF-OO / FIXC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X55450; GAA39093.1; -.  
 DR PIR; S14072; S14072.  
 KW Nitrogen fixation; Oxidoreductase; FAD; Flavoprotein.  
 KW NF\_BIND 12  
 FT SEQUENCE 435 AA; 48578 MW; A736998BDFAE34E2 CRC64;  
 SQ

Query Match 7.7%; Score 98.5; DB 1; Length 435;  
 Best Local Similarity 22.4%; Pred. No. 0.21;

Matches 52; Conservative 42; Mismatches 91; Indels 47; Gaps 8;

QY 7 EFINIVLSGGAAGIAHIGVAKINE--IGIRRALSGVSAAGIYVFAAGYSPBGMF 63  
 Db 209 EGVVLEAGTISRGTGNGFTYANKKESISLGI-----GCLVSDFOKTEYVGL- 257  
 QY 64 SLIKRVNMLKLFKPKPLKGLIGWEKAIKRFLEEVLP--YRRIEKLPIPIYICATDLYSG 120  
 Db 258 -----LEBKSHPSVAFLIEGSEVVEVAHLIPGCGKALPOLFGDGMVVVGAQQL 309  
 QY 121 RALYLSBSGLIPALIGSCAIPGIF-----EPVEYKRYLLVDDGIIVNNLPVEPPOESGIP 174  
 Db 310 NNAVREBSNLAMTSGRIAAEALITOVKSRBPMSAKKL-----SLKTYLSESFV- 359  
 QY 175 TVCVVULPIBEKXIKNITLILRSFLLAV-RNSSEKREKFCOLVIVPELAE 225  
 Db 360 -----LKQMKKKYKOLPALHINSQNFILTYPOLVSKAMQNFVAVDGPVKKEK 406

Search completed: June 19, 2003, 15:03:01  
 Job time : 23 secs

Thu Jun 19 17:30:47 2003

us-09-903-410-36.rpt

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 14:50:55 ; Search time 44 Seconds

(without alignments)  
1175,405 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273

Sequence: 1 LRKRFEINVLVSGAAGK.....RKAOIMRGVYKALVLS 251

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	934	73.4	259	16	067393
2	391.5	30.8	728	16	09HY6 aquifex seo
3	322.5	25.3	287	16	08R824
4	300.5	23.6	275	16	09K9Q8
5	297.5	23.4	764	16	09KUB9
6	294	23.1	319	16	08Y066
7	289	22.7	760	16	08R6F6
8	287.5	22.6	304	2	093KA7
9	285.5	22.4	491	2	093CN4
10	284.5	22.3	303	2	086196
11	284.5	22.3	321	16	092Q86
12	279	21.9	286	2	09RA24
13	276	21.7	293	16	08YB65
14	274.5	21.6	300	16	09JWV3
15	273	21.4	314	16	08YH96
16	273	21.4	315	16	08UEK6

17	268.5	21.1	297	16	09JXB2	09jxb2 neisseria m
18	268.5	21.1	395	16	09PH78	09ph78 xyella fas
19	261.5	20.5	301	16	08ZP44	08z44 salmonella
20	259	20.3	289	16	0983W7	0983w7 rhizobium l
21	258.5	20.3	314	16	08XDP1	08x41 escherichia
22	257.5	20.2	301	16	08Z7F2	08z7f2 salmonella
23	257	20.2	324	16	098LQ4	098lq4 rhizobium l
24	253.5	19.9	142	16	005884	005884 mycobacteri
25	251.5	19.8	289	2	09CRD8	09crd8 xanthomonas
26	248	19.5	295	16	092RQ6	092rq6 rhizobium m
27	247.5	19.4	1065	16	08VJ44	08vj44 mycobacteri
28	243	19.1	345	16	091385	091385 pseudomonas
29	240	18.9	311	16	083996	083996 treponema p
30	238	18.7	303	16	08Y2M1	08y2m1 raietonia s
31	235.5	18.5	610	16	09A6C3	09a6c3 caulobacter
32	233.5	18.3	466	11	08R064	08r064 mus musculu
33	233.5	18.3	1240	11	08R3C5	08r3c5 mus musculu
34	232.5	18.3	702	4	08YAY5	08yay5 homo sapien
35	232.5	18.3	409	4	096N75	096n75 homo sapien
36	231.5	18.2	311	16	08UHJ1	08uhj1 agrobacteri
37	231	18.1	1071	16	069695	069695 mycobacteri
38	228	17.9	905	5	08SVN8	08svn8 encephalico
39	223.5	17.6	1327	4	060859	060859 homo sapien
40	223.5	17.6	1327	11	09R114	09r114 mus musculu
41	208.5	16.4	1371	5	095PX1	095px1 caenorhabdi
42	199.5	15.7	296	16	09K945	09k945 bacillus ha
43	176.5	13.9	290	12	091F63	091f63 chilo iride
44	172.5	13.6	382	4	09UG58	09ug58 homo sapien
45	162	12.7	283	16	097GE3	097ge3 clostridium

# ALIGNMENTS

RESULT 1	ID	Score	Query Match	Length	ID	Description
067393	PRELIMINARY;	PRT;	259	AA.		
AC	067393	067393	067393	067393	067393	067393
DT	01-AUG-1998 (TREMURel. 07, Created)					
DT	01-AUG-1998 (TREMURel. 07, Last sequence update)					
DT	01-JUN-2002 (TREMURel. 21, Last annotation update)					
DB	Hypothetical protein A0_1386.					
GN	A0_1386.					
OS	Aquifex aeolicus.					
OC	Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.					
OX	NCBI_Taxid=63363;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=VFS;					
RX	MEDLINE=98196666; PubMed=9537320;					
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Jenox A.L.,					
RA	Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,					
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,					
RT	"The complete genome of the hyperthermophilic bacterium Aquifex					
RT	aeolicus."					
RL	Nature 392:353-358(1998).					
DR	EMBL; AB000737; AAC07357.1; -					
DR	InterPro; IPR002641; Patatin.					
DR	Pfam; PF01734; Patatin; 1.					
KW	Hypothetical protein; Complete proteome.					
SQ	SEQUENCE 259 AA; 29099 MW; 91C2B986255B6501 CRC64;					

Query Match	73.4%; Score 934; DB 16; Length 259;
Best Local Similarity	71.5%; Pred. No. 8.8e-77;
Matches 178; Conservative 31; Mismatches 40; Indels 0; Gaps 0;	
QY	1 LRKRFEINVLVSGAAGKIHIVKALNELGIRRAISGVSAGAIYVFAAGSYB 60
DB	5 LRKRFEINVLVSGAAGKIHIVKALNELGIRRAISGVSAGAIYVFAAGSYB 64
QY	61 GMSLLKRVNVLGLFKFKPLGLIGWEKAIKRLFEVLPYRLEKLEIPYICATLVYG 120
DB	65 EMLKTLKEVNVMLGLFKFKPKKGLIGWEKAEFLKEKELGVRLIEDLNIPIYLLSADLVYG 124

```

QY 121 RALVSEGLSIPALLGSCAIPFPEVVEYKYNLLVDGGIVNNLPVEPQESGIPVCVDV 180
DB 125 KALYFGRGDLFVLVSSCSIPGIFPEVYENFLVDGGIVNNLPVEPLEKKEPIGVVDV 184
QY 181 LPIEPEKDKINLHILRLSPFLAVRSNKRKFCFDLIVVPELEFPLDVRKADQIMER 240
DB 185 LPIEPEKDKINLHILRLSPFLAVRSNKRKFCFNNVTEPPELEFPLDVRKADQIMER 244
QY 241 GYIKALEVL 249
DB 245 GYESTLRIM 253

RESULT 2
Q9HYQ6 PRELIMINARY; PRT; 728 AA.
AC Q9HYQ6;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR ENBL; AB004756; AAG06727.1; -.
DR InterPro; IPR002641; Patatin.
DR Pfam; PF01734; Patatin; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 728 AA; 80898 MW; CF6A11B83D916043 CRC64;

Query Match 30.8%; Score 391.5; DB 16; Length 728;
Best Local Similarity 33.6%; Pred. No. 66-27;
Matches 95; Conservative 53; Mismatches 94; Indels 41; Gaps 6;

QY 8 EINVLSCGAAGKIAHIGLVKAINELGIRVALSGVSAGAIYVFYASGYSPGMPFLLK 67
DB 24 KIGLVSCGAAGKIAHIGLVKALDEQIGQIDAIAGTSMGAVVGGLYASGYTPAELEIAL 83
QY 68 RYNNLKLKFKKPLK-----GLIGWEKAIKRLPEE 96
DB 84 ENMDQALSDAPKDVFRKQDRDPLVKQKISFDGDTGLGLVQGNLAWLES 143
QY 97 VLPPY----RRIEKEIPIYICATLYSGRALYSEGLIPALGSCAIPGIPVEYKRY 152
DB 144 LLVHTSDNRDFDKLAIPRAVSTDIATGKVFVRKGLPQAIRASNSIPAVPAPEIDGR 203
QY 153 LLVDGGIVNNLPVEPQESGIPTV-CVDV-LPIEPEKDKINLHILRLSPFLAVRSNEK 210
DB 204 LLVDGGWMDNIPVDVARDMGVDVIVVDIGNFLDRKDLSTLVDVWNGITLMTKNSEA 263
QY 211 RKEFC---DLVIVPELEFPLDVRKADQIMERGYIKALEVLS 250
DB 264 QLATKPGDLVQLPPLSGYGTDRGVPQLIDAGY-RATTVLA 305

RESULT 3
Q8R824 PRELIMINARY; PRT; 287 AA.
AC Q8R824;
RA 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Predicted esterase of the alpha-beta hydrolase superfamily.
GN RSSA OR TTE1838.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB47 / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RA "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR ENBL; AB013136; AM25029.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 287 AA; 32028 MW; 80E698609ABCB34E CRC64;

Query Match 25.3%; Score 322.5; DB 16; Length 287;
Best Local Similarity 31.2%; Pred. No. 3.4e-21;
Matches 86; Conservative 48; Mismatches 107; Indels 35; Gaps 6;

QY 9 INVLSCGAAGKIAHIGLVKAINELGIRVALSGVSAGAIYVFYASGYSPGMPFLLK 68
DB 5 VGLILGGAGARGVAHLGILKRFEEINPIDFVIGISMGAIIIGAIYASGNNVDKLISSAKK 64
QY 69 VNNLK---LRFKPPKLGKIGWEKAIKRLPEEVLPIYRIEKEIPIYICATLYSGRALYL 125
DB 65 INMLKFTSLDDFKASRTGLVKGSEIKYLRGVK-ESFEELNIPLYIVATDIQTGKGVWF 123
QY 126 SEGSLIPALLGSCAIPGIPVEYKYNLLVDGGIVNNLPVEPQESGIPTV-----CV 178
DB 124 KEGDLIKAIRASISIPAFPEPVEYNGTKLVDSIVDSEAIELASGLADIINDCVSSI 183
QY 179 D-----VLPIEPEKDKINLHILRLSPFLAV---RSNKRKKEPCD--- 216
DB 184 DMGPFERFFYLSANSEKVLPIKRYFKIRPLPEISITTTAKLILKNSKNLEKMEGNK 243
QY 217 --LVIVPELEFPLDVRKADQIMERGYIKALEVLS 250
DB 244 RVTIKENVNIRYRFDQAEKCNMGPEADSIWA 279

RESULT 4
Q9K908 PRELIMINARY; PRT; 275 AA.
AC Q9K908;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein BH2587.
GN BH2587.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR ENBL; AP001516; BAB06306.1; -.

```

DR InterPro: IPR002641; Patatin.  
 DR InterPro: IPR001423; UPF0028.  
 DR Pfam: PF01734; Patatin; 1.  
 DR PROSITE: PS01237; UPF0028; 1.  
 DR Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 275 AA; 30545 MW; 617C790E9AE553 CRC64;

Query Match 23.6%; Score 300.5; DB 16; Length 275;  
 Best Local Similarity 31.4%; Pred. No. 3.2e-19;  
 Matches 71; Conservative 48; Mismatches 100; Indels 7; Gaps 3;

QY 8 EYNLVSGGAAGKAIHIGVKAINEIGIRVRLSGVSAAGIYSVFYASGSPGEMF---S 64  
 DB 8 KIGLVSGGARGAHGIVLKVLEQEKIPIDYLAGSSMALVSLYSGHGTETELIRAN 67  
 QY 65 LKRVNMLKFKPKPLKGLIGMEKAIPLLEVLPRRIEKLPIYICATDLYSGALY 124  
 DB 68 LFKRYKYL---DFVVPKRGFLAGHRYVELIRYLAKKKRYEELDPVRIVAADLTKGRVI 124  
 QY 125 LSEGSLLIPALGSCAIPGIFEPVEYKNYLLVDGIVNNLPVPEPQESGIP--TVCVDVLP 183  
 DB 125 LQEGVAVAVRASIAIGIFVPKNINDRLIDGGVIDRVPVGVVEMGADLTIVADLTIF 184  
 QY 184 EEEKIKNIHLILRSFPLAVRSSEKKEFCDLVVPLEEFPTL 229  
 DB 185 REELIRSVYVILQITDMMSKELVRVQIDCTVMIRPNDRYSL 230

## RESULT 5

Q9KUB9 PRELIMINARY; PRT; 764 AA.

AC 09KUB9: 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical protein VCO603.  
 GN VCO603.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heideberg J.F., Bisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ertolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Niemann W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.D., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nucleic Acids 406:477-483(2000).  
 DR EMBL: A8004145; AAF93770.1; --  
 DR TIGR: VC0603; --  
 DR InterPro: IPR002106; AAcRNA\_1igaseII.  
 DR InterPro: IPR002641; Patatin.  
 DR Pfam: PF01734; Patatin; 1.  
 DR PROSITE: PS00339; AA TRNA LIGASE II 2; UNKNOWN\_1.  
 DR Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 764 AA; 85616 MW; 0C3C8BCA100C840 CRC64;

Query Match 23.4%; Score 297.5; DB 16; Length 764;  
 Best Local Similarity 27.6%; Pred. No. 2.2e-18;  
 Matches 79; Conservative 61; Mismatches 105; Indels 41; Gaps 7;

QY 2 RLKREELNVLVSGGARGAHGIVLKAINEIGIRVRLSGVSAAGIYSVFYASGSPG 61  
 DB 28 QVAKRPEKILVLAGGARGAHGIVLKAINEIGIRVRLSGVSAAGIYSVFYASGSPG 87  
 QY 62 MSLKRVNMLKFK-----FKPPLKGLIGMEKA 90

DB 88 IEALIVSDMNGVRDVRDVSQRVRDEKYEEDRYQITLTLGLHKEVRAP--KGVVQGNM 146  
 QY 91 IRLRELV---LP-YRIKELPIPTICATDLYSGRALYLSGSLIPALGSCAIGIEP 146  
 DB 147 LRLRLRTTNLRAFSFQVLVYRAVATDILHQLQEVLEKGLVDMAMASVFGALP 206  
 QY 147 VEYKNYLLVDGIVNNLPVPEPQESGIP--TVCVDV--LPLEPKDNIHLILRSFPLAV 204  
 DB 207 YEIDGMLVDGVTNNMVEVARMGADIIIVADISTDYSGDEFTNLTVAQLSNLY 266  
 QY 205 RSNSEKKEKFC---DEVIVPELEEFPTLVDVAKADQIMERGIVALE 247  
 DB 267 RSTERSDHLTSRLDLRFPVGKMETEPKMPARAPMAYQGRMD 312

## RESULT 6

Q8Y066 PRELIMINARY; PRT; 319 AA.

AC 08Y066: 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Probable lipoprotein transmembrane.  
 GN RSC1178 OR RS04543.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum) 247  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OX Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Airlant M., Billault A., Brothier P., Camus J.C., Catolico L.,  
 RA Chander M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nucleic Acids 415:497-502(2002).  
 DR EMBL: A646063; CAD14880.1; --  
 DR InterPro: IPR002641; Patatin.  
 DR Pfam: PF01734; Patatin; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 319 AA; 32967 MW; 6924E744D4F53342 CRC64;

Query Match 23.1%; Score 294; DB 16; Length 319;  
 Best Local Similarity 32.4%; Pred. No. 1.5e-18;  
 Matches 84; Conservative 51; Mismatches 100; Indels 24; Gaps 8;

QY 9 INVLVSGGAAGKAIHIGVKAINEIGIRVRLSGVSAAGIYSVFYASGSP--EGMFSLLK 67  
 DB 53 IGLALGGAARGAHGIVLKAINEIGIRVRLSGVSAAGIYSVFYASGSP--EGMFSLLK 112  
 QY 68 R-----VNNLKLF--KPKPLKGLIGMEKAIPLLEVLPRRIEKLPIYICATDLYSG 120  
 DB 113 MDEATLADMLPFGTFFGMLG---BALRKYNRLVKOKTIRBWKIPILGIVADLSSG 168  
 QY 121 RALYLSGSLIPALGSCAIPGIFEPVEYKNYLLVDGIVNNLPVPEPQESGIPVCDV 180  
 DB 169 RPIILFRGNTGVVRASGIPGVFPVITSGHQVVDGVLVAVPVYAKMGATFVIAVN 228  
 QY 181 LPLEPE-KDKIKNIHLILRSFPLAVRSSEKKEFCDLVVPLEEFPTLVDVAKADQIMERGIVALE 234  
 DB 229 ISADPSKQAVSGASMLDPTTITMOSINKIELAQADVIAFSLPFGVGSDF---ARN 284  
 QY 235 DQIM--ERGVKALEVLS 251  
 DB 285 EALVGEQALAMPLIRE 303

## RESULT 7

Q8R6F6





```

QY 206 SNSKRRKPCDLYVPELEETPLDVRKADQIMERGYTAKALEVL 249
Db 264 KTDYSLPDSAGILMTFTKXNDVSLMDFORIDELKIGYDRTMSLM 307

RESULT 10
ID 086196 PRELIMINARY; PRT; 303 AA.
AC 086196;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ExpL protein.
GN Pectobacterium carotovorum subsp. carotovorum.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
CX NCBI_TaxID=555;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RA Andersson R.A., Palva T.E., Pirhonen M.;
RT "The response regulator ExpM is essential for the virulence of Erwinia
RT carotovora subsp. carotovora and acts negatively on the sigma factor
RT Rpos."
RL Mol. Plant Microbe Interact. 0:0-0(0).
DR EMBL, AJ224437; CAU1941.1; -.
DR InterPro; IPR002641; Patatin.
DR Pfam; PF01734; Patatin; 1.
SQ SEQUENCE 303 AA; 33406 MW; FE972380ACAD1F9 CRC64;

Query Match 22.3%; Score 284.5; DB 2; Length 303;
Best Local Similarity 28.3%; Pred. No. 1e-17; Indels 61; Gaps 7;
Matches 82; Conservative 46; Mismatches 101;

QY 7 BEINIVLSGGAAGIAHIGVLAINELGIRVALSGVSAIVSVFYASGYSPGEMFSL 66
Db 4 KIGIVLSGGAAGIAHIGVLAINELGIRVALSGVSAIVSVFYASGYSPGEMFSL 66
QY 67 KRVNMLKLPFKRPPK-----GLIGWEKAIPLFEVLPYRIEKLIEPTTCARDL 117
Db 57 SMEQWVRFGEYWDVIRLMDLMSQSGILRGDRVFNVSVKLTHTTQIEDCAIKYGVVTNL 116
QY 118 YSGRALYISEGSLIPALLGSCAIPGIFEEVEYKNVLLVDGIVNNLPVEPFQESGIPTVC 177
Db 117 STGRELMTEGDLHQMRASCMPGLSPFRFNDYLVDAVAVNPVPSLAPAMGADIVI 176
QY 178 V-----DVLPIEP-----EKDI-----KNILHLRL----- 198
Db 177 AUDLOHDASINMODLSIKETVSDVIDMEHVSQDMSRIIRRELLRGRQASBPPTAME 236
QY 199 --SFLVANSSEKKEKPC---DLVIVPELEETPLDVRKADQIMERGY 242
Db 237 IMSTSIQIENRLKMTRMAGDPDPVILQPYCQIATLDHRAQEAIEAGY 286

RESULT 11
ID 092086 PRELIMINARY; PRT; 321 AA.
AC 092086;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein K01418.
GN R01418 OR SMC01003.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
CX NCBI_TaxID=382;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=1021;

```

```

RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boizard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ransperger U.,
RA Renard C., Thebaud P., Vandendol M., Weidner S., Gallibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL, AF591787; CAC45997.1; -.
DR InterPro; IPR002641; Patatin.
DR Pfam; PF01734; Patatin; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 321 AA; 34380 MW; D9AD7590B84D709F CRC64;

Query Match 22.3%; Score 284.5; DB 16; Length 321;
Best Local Similarity 30.4%; Pred. No. 1e-17; Indels 51; Gaps 6;
Matches 86; Conservative 47; Mismatches 99;

QY 9 INIVLSGGAAGIAHIGVLAINELGIRVALSGVSAIVSVFYASGYSPGEMFSL 64
Db 40 IALALGGAARGWAIHIGVLAINELGIRVALSGVSAIVSVFYASGYSPGEMFSL 99
QY 65 LKRVNMLKLPFKRPPKLGIGWEKAIPLFEVLPYRIEKLIEPTTCARDL 124
Db 100 TVRRI--AGILDPALGGGIFGGLRLTKMQEHLQMLSTEDLPFVAVATEVYSGHEW 157
QY 125 LSEGSLIPALLGSCAIPGIFEEVEYKNVLLVDGIVNNLPVEPFQESGIPTVC-----VD 179
Db 158 IEKSLITAIRASYALPGIFEPNAGRTLIDALVNPV-----SVCRARHQH 208
QY 180 VLPPIEPKDI-----KNILHLRSPFLVANSSEK 210
Db 209 VVANINLVYVGRSAVYKSGMETPDAPATANRESARVIGMTSVWQAFNIQDRISBA 268
QY 211 R--KEPCDLYVPELEETPLDVRKADQIMERGYTAKALEVL 251
Db 269 RLADPDPDLIHPRLNDIGLSEPHRAGEALIERGOEAKTYLSE 311

RESULT 12
ID 09RA24 PRELIMINARY; PRT; 286 AA.
AC 09RA24;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Genes, similar to eicosapentaenoic acid synthetase gene cluster.
OS Vibrio marinus (Moritella marina).
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Moritella.
CX NCBI_TaxID=90736;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MP-1;
RA Tanaka M., Ueno A., Kawasaki K., Yumoto I., Ohgiya S., Hoshino T.,
RA Ishizaki K., Okuyama H., Morita N.;
RT "Isolation of clustered genes that are notably homologous to the
RT eicosapentaenoic acid biosynthesis gene cluster from the
RT docosahexaenoic acid-producing bacterium Vibrio marinus strain MP-1."
RL Biotechnol. Lett. 21:939-945(1999).
DR EMBL, AB025342; BAA69379.1; -.
DR InterPro; IPR002641; Patatin.
DR Pfam; PF01734; Patatin; 1.
SQ SEQUENCE 286 AA; 31587 MW; AAF729C1D2CEB6C9 CRC64;

Query Match 21.9%; Score 279; DB 2; Length 286;
Best Local Similarity 32.9%; Pred. No. 3e-17; Indels 12; Gaps 5;
Matches 80; Conservative 42; Mismatches 109;

QY 8 BEINIVLSGGAAGIAHIGVLAINELGIRVALSGVSAIVSVFYASGYSPGEMFSL 64
Db 7 KIGIVLSGGAAGIAHIGVLAINELGIRVALSGVSAIVSVFYASGYSPGEMFSL 66

```

```

QY 65 LLKRVNWLKLFKPKPLKGLIGWEKAIHFLFEEVLPYRIEKLPIPTVICATDLYSGRALY 124
Db 67 DVKPSW-----KFTPARAGFDPAKLYPEVLKYIPEDSPFYQLPELRIVATNWLKKEHI 122
QY 125 LSEGLIPALGSCAIPGIFRVEYKYNLLVDGIVNNLVPFQSGIPTVCVDVLPTE 184
Db 123 FKDGSVINALLASAYPLVFPFMIIDDQVYSDGIVNHFVPSVIEDCDKIIGVYSPIR 182
QY 185 PEK--DIKNILHILSRPFLAVRNSSEKKEFCFLVIVPE--LSEFTPL--DVKRADOIME 239
Db 183 QVEADELSIKDVLRAFTLQSGNELDKLSQCDVQIYPEALLNYNTPATDEKSLREIVQ 242
QY 240 RGY 242
Db 243 IGY 245

RESULT 13
ID QYB65 PRELIMINARY; PRT; 293 AA.
AC QYB65;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Serine protease (EC 3.4.21.-).
GN BMEI11035.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapattal V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haseelkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL; AL54277.1; -.
DR InterPro: IPR002641; Patatin.
DR Pfam: PF01734; Patatin; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 293 AA; 31838 MW; D90C3455DF539474 CRC64;

Query Match 21.7%; Score 276; DB 16; Length 293;
Best Local Similarity 28.5%; Pred. NO. 5.8e-17;
Matches 73; Conservative 52; Mismatches 99; Indels 32; Gaps 7;

QY 11 LVLGGGAAGIAHIGVLAINEGIRVRALSGVAGATVSFVYASGYSPEGMFSLKRV- 69
Db 13 VAFGGGARGIAHIVVDVNLGLQPVIAAGSIGSIGSSMANGSKDIYDYNASIF 72
QY 70 -----NWLKLFKPKPLKGL-----IGWEKAIHFLFEEVLPYRIEKLPI 108
Db 73 TRNSRVVRMWTQPAHNAELPK-----GGLRVSPNIEKYLDVFLPDSPF-SNVEDLKI 126
QY 109 PTVICATDLYSGRALYLSGSLIPALLGSCAIPGIFRVEYKYNLLVDGIVNNLVPF 168
Db 127 PTTITAVDFEGAREIHISDGLRSALASCAIPVPFAPVRNGRILVDGGLFNPFPDLL 186
QY 169 QESGIPVCVDVL--PIEFKDIKNILHILSRPFLAVRNSSEKKEF--CDLVIVPELE 224
Db 187 FDKADIVIGDVGVGVPDDEMTPTFEAVIGTSQLTMCSTIENKFRYPHPHIFVRPNVE 246
QY 225 EFTPLDVRKADQIMER 240
Db 247 RIGLLDFKFEQVISQ 262

```

```

RESULT 14
ID Q9JWV3 PRELIMINARY; PRT; 300 AA.
AC Q9JWV3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Possible lipoprotein.
GN NMA0225.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506 (2000).
DR EMBL; AL162752; CAB83536.1; -.
DR InterPro: IPR002641; Patatin.
DR Pfam: PF01734; Patatin; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 300 AA; 31367 MW; 8CF371BE0F5F51CF CRC64;

Query Match 21.6%; Score 274.5; DB 16; Length 300;
Best Local Similarity 31.8%; Pred. No. 8.2e-17;
Matches 71; Conservative 40; Mismatches 97; Indels 15; Gaps 3;

QY 9 INLVSGGAAGIAHIGVLAINEGIRVRALSGVAGATVSFVYASGYSPEGMFSLKRV- 68
Db 46 VGLALGGGARGIAHIGVLAINEGIRVRALSGVAGATVSFVYASGYSPEGMFSLKRV- 105
QY 69 VNMWKLKFKPKPLKGLIGWEKAIHFLFEEVLPYRIEKLPIPTVICATDLYSGRALYSEG 128
Db 106 LKGTDLVDLTLTSTSGFIKGEKQNYINRVKGGRIQQFPFKFAAVATDFETGKAVAFNOG 165
QY 129 SLIPALLGSCAIPGIFRVEYKYNLLVDGIVNNLVPFQSGIPTVCVDVLPTEPEKD 188
Db 166 NAGQAVRASALIPNVFPVPIGRHTYVDGGLSQPVFVSARRQGANFVIADVISPSPKN 225
QY 189 IKNIHLILSRP-----LAVRNSSEKKEF--CDLVIVPEL 223
Db 226 IS-----QGPFSLDQTLNMSVSALQNELGQADVVIKPOV 261

RESULT 15
ID Q8YH96 PRELIMINARY; PRT; 314 AA.
AC Q8YH96;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Serine protease (EC 3.4.21.-).
GN BMEI0908.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapattal V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,

```

RA Selkov B., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
RA Haselkorn R., Kyriides N., Overbeek R.,  
RT "The genome sequence of the facultative intracellular pathogen  
RT Brucella melitensis." U.S.A. 99:443-448(2002).  
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
DR EMBL; AE009531; AAL52089.1; -.  
DR InterPro; IPR002641; Patatin.  
DR Pfam; PF01734; Patatin; 1.  
KM Hydrolase; Complete proteome.  
SQ SEQUENCE 314 AA; 34266 MW; E19ADC1225AFDB CRC64;

Query Match 21.4%; Score 273; DB 16; Length 314;  
Best Local Similarity 27.8%; Pred No. 1.2e-16;  
Matches 80; Conservative 52; Mismatches 98; Indels 58; Gaps 7;  
QY 7 BEINLVISGAGAKGIMHIGVLKAINELGIRVALSGVSAGATVSFYAGYSPE----- 60  
DB 33 QKIALALGGGAAGMWHIGVLRALDAGIEIEMIAGTISIGALVGGCYLAGKNELEEFAR 92  
QY 61 -----GWFSLKKEVNMKLKFKPPPLKGLIGMEKAIKFLSEVLPPRIEKLBIPTYICAT 115  
DB 93 SLTRRRMFMNLD-----ITPRG--SGLFQGMKLDGRLREHLDGIRIEDLPKPFVAVCT 143  
QY 116 DLYSGRALYISBSGLIPALSGCAIPGFEEVEYKNTLVNDDGIVNNLPVE----- 166  
DB 144 ELKTEHETWLTSTGPLVEARASYALPGVFEPRMGDRVLVDGALVNPVFPVSCRAYEORL 203  
QY 167 -----PFQESGIPTVCCVDVLPFBENDIK-NILHILRSFFLAVR 205  
DB 204 VLAANLHYDQGRAAVIKQAKRORSVPOA-----MHGSEKTRLGITGVMEAFNIID 257  
QY 206 SNSEKPK--EPCDLVIVPELEBFTPLDVRKADQIMERGKYLKALEVLSB 251  
DB 258 RISRAARAGDPPDVSLMPTVGQIGLADFHRAAALDAGYIETVARIED 305

Search completed: June 19, 2003, 15:04:02  
Job time : 45 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 12:47:00 ; Search time 173 Seconds  
(without alignments)  
9841.102 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756  
Sequence: 1 ttcgagattcgaggaatttga.....tagagctacttcgtag 756

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	752.8	99.6	756	18	AAT79324
2	396.2	52.4	750	18	AAT79327
3	53.6	7.1	903	20	AAZ12214
4	52	6.9	903	20	AAZ12213
5	52	6.9	92934	21	AAA81473
6	52	6.9	172325	21	AAAF21613
7	52	6.9	837096	21	AAA81489
8	49.8	6.6	1877	23	AAAT7947
9	47.4	6.3	903	20	AAZ12215

10	46.2	6.1	40138	20	AAV81946	V. marinus PK8-lik
11	46.2	6.1	40138	21	AAA71520	V. marinus PK8-lik
12	46.2	6.1	41587	22	AA164984	Moritella marina i
13	45.8	5.1	467	24	ABK79457	Bacillus clausii g
14	42.6	5.6	447	20	AAZ12212	Neisseria meningit
15	42.6	5.6	447	21	AAAB1390	N. meningitidis Me
16	39.4	5.2	269223	22	AAF28554	Genomic fragment #
17	39.4	5.2	12019	20	AAAX1867	Alcaligenes sp. Po
18	36.6	4.8	1497	22	AAAF1015	P. putida KT2440-a
19	35.8	4.7	876	24	ABK75059	Bacillus lichenifo
20	35.4	4.7	5591	23	ABL04293	Drosophila melanog
21	35.4	4.7	13437	23	ABL04292	Drosophila melanog
22	35.4	4.6	1484	21	AAAC4712	Arabidopsis thalia
23	35	4.6	1487	21	AAAC36225	Arabidopsis thalia
24	34.4	4.6	11628	20	AAAX20520	Polynucleotide seq
25	34	4.5	1299	20	AAAX20681	Polynucleotide seq
26	33.2	4.4	5520	19	AAV03311	Thermococcus 9N-2
27	33	4.4	472	22	AAH84381	E. coli growth and
28	33	4.4	10944	22	AAH46250	DNA encoding novel
29	32.8	4.3	3793	24	ABK92234	Prostate cancer-as
30	32.8	4.3	5948	23	ABL14817	Drosophila melanog
31	32.8	4.3	11046	23	ABL14816	Drosophila melanog
32	32.8	4.3	34980	22	AAH41223	Pyrococcus abyssi
33	32.6	4.3	34980	22	AAH41224	Pyrococcus abyssi
34	32.6	4.3	494	22	AAI36495	Probe #5181 used t
35	32.6	4.3	494	22	ABK505206	Human genome-deriv
36	32.6	4.3	567	24	ABK53343	Human eosinophil-m
37	32.4	4.3	994	24	ABQ69148	listeria monocytog
38	32.4	4.3	994	24	ABQ70659	listeria monocytog
39	32.4	4.3	2944528	24	ABA303041	Listeria monocytog
40	32.2	4.3	4403765	22	AAI99683	Mycobacterium tube
41	32.2	4.3	4411529	22	AAI99682	Mycobacterium tube
42	32	4.2	811	24	AAQ89282	Human prostate exp
43	32	4.2	1476	21	AAAC47228	Arabidopsis thalia
44	32	4.2	1478	21	AAAC42167	Arabidopsis thalia
45	32	4.2	17569	24	ABK39756	cDNA encoding clon

## ALIGNMENTS

RESULT 1  
AAT79324 standard; DNA; 756 BP.

AC AAT79324;

DT 16-FEB-1998 (first entry)

DE DNA encoding Aquifex pyrophilus esterase 28LC.

EW Esterase; thermostable enzyme; ester; chiral compound; cheese;

KW pulp; paper; lignin removal; sugar; lignocellulose;

KW disease resistance; feedstuff; ss.

OS Aquifex pyrophilus strain KO1 5a.

FT Key Location/Qualifiers  
FT CDS 1..756 /\*tag= a  
FT FT /transl\_except= (pos: 742..746, aa:Val)

PN WO9730160-A1.

PD 21-AUG-1997.

PF 11-FEB-1997; 97WO-US02039.

PR 16-FEB-1996; 96US-0602359.

PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

PI Callen W, Kosmicka A, Link S, Maffia AM, Murphy D;

Reid J, Robertson DE, Swanson RV, Warren PV;  
WPI; 1997-425035/39.  
P-PSDB; AAW23071.

Nucleic acid encoding heat stable esterase from thermophilic bacteria - which is active in organic solvents, useful in cheese or paper manufacture, and to study plant resistance to disease

Claim 1; Page 44-45; 113pp; English.

This DNA sequence codes for thermostable esterase 28LC (AAW23071) of Aquifex pyrophilus, a Gram-negative, chemolithoautotrophic knall gas marine bacterium which grows optimally at 85 deg C and pH 6.8. It can be amplified from a pBluescript vector by PCR (see AAT79306-07). Claimed, newly identified polynucleotides (AAT79321-30) encoding esterases (AAW23069-77, AAW23088) were recovered from genomic gene libraries. They can be used for recombinant production of the enzymes in host cells, and as probes to identify related sequences. The esterases are stable at high temperature and in organic solvents, making them superior for use in pharmaceutical, agricultural and other chemical industries. A method is claimed for transferring an amino group from an amino acid to an alpha-keto acid using a claimed esterase. The enzymes are stable at high temperature and in organic solvents, making them superior for use in production of pure chiral compounds used in pharmaceutical, agricultural and other chemical industries. A method is claimed for transferring an amino group from an amino acid to an alpha-keto acid using a claimed esterase. The enzymes

Query Match 99.6%; Score 752.8; DB 18; Length 756;  
Best Local Similarity 99.7%; Pred. No. 3e-239;  
Matches 754; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TTGAGATTGAGAAATTTGAAGAGATAAACCCTGCTTTTCGGAGGAGCTGCAAGGGC 60  
1 TTGAGATTGAGAAATTTGAAGAGATAAACCCTGCTTTTCGGAGGAGCTGCAAGGGC 60

61 ATAGCCCATAGAGTGTGTTTGAAGAGCTATAACGAGCTCGGTATTAAGGGTGAGGCTTTA 120  
61 ATAGCCCATAGAGTGTGTTTGAAGAGCTATAACGAGCTCGGTATTAAGGGTGAGGCTTTA 120

121 AGCGGGTGAGCGCGCGGGCAATCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAA 180  
121 AGCGGGTGAGCGCGCGGGCAATCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAA 180

181 GGGATGTTTTCAGCTTCTGAAAGAGGGTAACTGGCTGAAAGCTGTTTAAAGTTCAAGCCACT 240  
181 GGGATGTTTTCAGCTTCTGAAAGAGGGTAACTGGCTGAAAGCTGTTTAAAGTTCAAGCCACT 240

241 CTGAAGGATTTGATAGGTTGGGAGAGGCTATAAGATCTCTTGAGGAAAGTTCTCCCTTAC 300  
241 CTGAAGGATTTGATAGGTTGGGAGAGGCTATAAGATCTCTTGAGGAAAGTTCTCCCTTAC 300

301 AGAGAAATAGAAAACCTTGAGATACCGACGTATATATGCGGACGAGTTTATATATCGGGA 360  
301 AGAGAAATAGAAAACCTTGAGATACCGACGTATATATGCGGACGAGTTTATATATCGGGA 360

361 AGGGCTCTATACCTCTCGAAGGAGTGTAAATCCCGCACTTCTCGGACAGCTGTGCAATT 420  
361 AGGGCTCTATACCTCTCGAAGGAGTGTAAATCCCGCACTTCTCGGACAGCTGTGCAATT 420

421 CCCGGCATTTTGAACCCGTTGAGTATTAAGAAATTAATCTGCTGTTGAACGAGGTATAGTT 480  
421 CCCGGCATTTTGAACCCGTTGAGTATTAAGAAATTAATCTGCTGCTGTTGACGAGGTATAGTT 480

481 AACAACTTCCCGTTGAGCCCTTTCAGGAAGCGGTATTTCCACACCGTTTGGCTTATGTC 540  
481 AACAACTTCCCGTTGAGCCCTTTCAGGAAGCGGTATTTCCACACCGTTTGGCTTATGTC 540

541 CTTCCATAGAGCCGGAAGAGATATATAAGAAATTTCTTACATCCTTTTGAGGAGCTTC 600  
541 CTTCCATAGAGCCGGAAGAGATATATAAGAAATTTCTTACATCCTTTTGAGGAGCTTC 600

601 TTTCTTGGCTCCGCTCAAACTCCGAAAGAGAGAGTGTGTTTGACCTCGTTATAGTT 660  
601 TTTCTTGGCTCCGCTCAAACTCCGAAAGAGAGAGTGTGTTTGACCTCGTTATAGTT 660

661 CTTGAGCTTTGAGGAGTTTCAACCCCTTCATGTTTAAAAAGCGGACCAATTAATGAGAGG 720  
661 CTTGAGCTTTGAGGAGTTTCAACCCCTTCATGTTTAAAAAGCGGACCAATTAATGAGAGG 720

721 GGATACATTAAGGCTTTAGAGTACTTTCTGAATAG 756  
721 GGATACATTAAGGCTTTAGAGTACTTTCTGAATAG 756

RESULT 2  
AMT79327  
ID AAT79327 standard; DNA; 750 BP.  
XX  
AC AAT79327;  
XX  
DT 16-FEB-1998 (first entry)  
XX  
DE DNA encoding Aquifex esterase VF5-34LC.  
XX  
KW Esterase; thermostable enzyme; ester; chiral compound; cheese;  
KW pulp; paper; lignin removal; sugar; lignocellulose;  
KW disease resistance; feedstuff; ss.  
XX  
OS Aquifex sp. strain VF5.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..750  
FT FT /\*tag= a  
FT FT /transl\_except= (pos: 295..297, aa:Glu)  
XX  
XX WO9730160-A1.  
XX  
XX 21-AUG-1997.  
XX  
XX 11-FEB-1997; 97WO-US02039.  
XX  
XX 16-FEB-1996; 96US-0602359.  
XX  
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.  
XX  
XX Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;  
XX Reid J, Robertson DE, Swanson RV, Warren PV;  
XX  
XX WPI; 1997-425035/39.  
XX  
XX P-PSDB; AAW23071.  
XX  
XX Nucleic acid encoding heat stable esterase from thermophilic  
XX bacteria - which is active in organic solvents, useful in cheese or  
XX paper manufacture, and to study plant resistance to disease  
XX  
XX Claim 1; Page 48-49; 113pp; English.  
XX  
XX This DNA sequence codes for thermostable esterase VF5-23LC  
XX (AAW23074) of Aquifex VF5, a marine strictly chemolithoautotrophic  
XX knall gas bacterium that grows optimally at 85 deg C and pH 6.8.  
XX It can be amplified from a pBluescript vector by PCR (see AAT79312-  
XX 13). Claimed, newly identified polynucleotides (AAT79321-30) encoding  
XX esterases (AAW23069-77, AAW23088) were recovered from genomic libraries.  
XX They can be used for recombinant production of the enzymes in host  
XX cells, and as probes to identify related sequences. The esterases  
XX are stable at high temperature and in organic solvents, making them  
XX superior for use in production of pure chiral compounds used in  
XX pharmaceutical, agricultural and other chemical industries. A  
XX method is claimed for transferring an amino group from an amino  
XX acid to an alpha-keto acid using a claimed esterase. The enzymes

CC may also be useful as ripening starters in cheese making, in lignin  
CC removal in paper and pulp manufacture, in carbohydrate derivative  
CC synthesis, in fermentable sugar production from lignocellulosic  
waste, in the study of plant wall structure, plant resistance to  
CC disease and organic matter decomposition and to select plants bred  
CC for production of highly degradable animal feeds.

**SQ** Sequence 750 BP; 214 A; 142 C; 193 G; 201 T; 0 other;

Query match 52.4%; Score 396.2; DB 18; Length 750;

Best Local Similarity 72.2%; Pred. No. 7, 9e-12i;  
Matches 515; Conservative 0; Mismatches 198; Indels 0; Gaps 0

QY	1	TTGAGATTGAGGAAATTGAGAGATTAACCTCGTTCTTTGCGGAGGAGCTGCAAGGGC	60
Db	13	TTGAAATTGAGGAGTTTGAAGGTTAACTTAATTCTTTCCGGAAGGGGCTGCAAGGGT	72
QY	61	ATAGCCCAATAGTGTGTTTGAAGACTTAAACGAGCTCGGTATTAAGGTGAGGGCTTTA	120
Db	73	ATGCCCCAATAGTGTGTTTAAACCTCTGGAAGAGCTCGGTATTAAGGTGAAGGCTC	132
QY	121	AGCGGGGTGAGCGCGGGGCATCGTTTCGGTCTTTATGCTCAGGCTACTCCCTGAA	180
Db	133	AGCGGGGTGAGGTCTGAGCTATCGTTTCGGTCTTTATGCTCAGGCTACTCCCTGAA	192
QY	181	GGGATGTTGAGCTTCTGGAAGGGTAAACTGCTGAGAGCTGTGTTAAAGTCAAGCACC	240
Db	193	GAGATGTAAATCTCTGAAAGAGGTAAACGCGCTCAAACTTTTAAGTTCAAAACACCG	252
QY	241	CTGAAGGATTTAGGTAGGGTGGAGAAAGGCTATAGATTCTTGAAGAGTTCTCCCTAC	300
Db	253	AAATATGGCTTAATATGGGGTGGAGAAAGGTGACAGTTTGTAAAAAGAGCTCGAGATT	312
QY	301	AGGAGATATGAAAACTTGAGATACCGACGTATATATCGCGACGAGATTATCTCGGGA	360
Db	313	AAGAGCTGAGAGACTGAACATTAACACCTATCTTCTGCGGAGATCTGTACAGGGA	372
QY	361	AGGACTCTATACCTCTCGAGAGGAGATTATCCCGACATCTCGGACGTGCAATT	420
Db	373	AAGGCTTTTACTTGGGCAAGGTGACTTAATCCCGGCTTCGGAAGTGTGCATAT	432
QY	421	CCGCGCATATTTGAACCCCGTTGAGTATTAAGATTCTTGCTCGTTGACGAGGTATAGT	480
Db	433	CCCGGATTTTGTAAACAGTGTAGTACGGAATTTTCACTGTGTGACGAGGTATAGTG	492
QY	481	AACAACCTTCGCTGAGGCCCTTTCAGGAACCGGATATCCCAACGTTTGCGTGAATGC	540
Db	493	AACAACCTTCGCGGAGAACCTTTGGAAGATTCAAGAACCAATATCGGGGTATGATGTG	552
QY	541	CTTCCCATPAGAGCGCGAAAAAGATATTAAGAACATTTCTTCACATCTTTTGAAGGCTTC	600
Db	553	CTTCCCATTACTCAAGAAAGAAAGATTAAAAATATATCTCAATCTTATPAGAGGCTTC	612
QY	601	TTTCTTGCGGTCCGCTCAAACTCCGAAAAAGAAAGAGTTTGTGACCTCGTATATGTT	660
Db	613	TTTCTGCGCGGTGCTTCAATTCGAAAAAGAAAGAGTTCTGCAAGTATGATTAGAA	672
QY	661	CCTGAGCTTGAAGAGTTCAACCCCTGTAGTGTAGAAAAAGGGAACCAATAT	713
Db	673	CTTCCCTCTGAAGGTTCTCTCTCTGACGTAAATAGCGGAGAGATATTT	725
RESULT 3			
AAZ12214			
ID	AAZ12214 standard; DNA, 903 BP.		
XX	AAZ12214;		
XX	08-OCT-1999 (first entry)		
DT	Neisseria meningitidis strain A complete ORF137 sequence.		
XX	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		

KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea; ss  
XX  
OS *Neisseria meningitidis*.

Query Match	Similarity	7.1%	Score 53.6	DB 20	Length 903
Best Local	Similarity	4.5%	Pred. N5.2e-07		
Matches	212	Conservative	0	Mismatches	264
				Indels	0
				Gaps	0
QY	19	GAGAGATPAACCTCGTCTTCTTCGGGAGAGAGCTGAAAGGGCATN3ACCCACATAGTGT	78		
DB	130	GGAGGGTCGGTTTGTGGCACTGGTGCGCGCGCATTTAAAGGATTTGCCCATGTAGTATT	189		
QY	79	TGAAGACTATPAACGAGCTCGGTATTAAGGGTGAGGGCTTTTAAGCGGGGTGAGCCCGGG	138		
DB	199	ATTAAAGTTTGAAGAAAGAGGTATTCGTGAAGAGTGGTTACCGGCACATCGCAGGT	249		
QY	139	GCAATCGTTTCGGTCTTTTATAGCCTCAGGCTACGCCCTCGAAGGATGTTCAGCCTTCG	198		
DB	250	TGGAATGTCGGAGCGCTTTTGGCGAGGTATGTGCGCCGACCGCTCGAATTGGAGCC	309		
QY	199	AAGAGGGTAACTGCGCTGAAAGCTGTTTAAGTTCAGGCCACCTGTGAAGGATTTGATAGG	258		
DB	310	GAAATTTAGTTPAAGCAATTTGGTTCATTTAACTTGTGCACCAAGTGTATTATCAA	369		
QY	259	TGGGGAAGCGTATTAAGTTCCTGTAGAGAAGTTCCTCCCTTACAGGAGATTAAGAAAATT	318		
DB	370	GGCGAAGAACTGCAAAATTTACATCAACCGAAAGTGGGGGAGCGGATTTGACAGATT	429		
QY	319	GAGATACCGACGTATATATATGCGCAGAGATTATATCTGGGAAGGGCTCTATACCTTCG	378		
DB	430	CCCATCAAAATTTGGCGCGCTTGTCTACTGATTTTGAAGCCGGGAAGGCCCTGCTTTCAT	489		
QY	379	GAAAGGATTTATCCCCGCACTTCTCGGCACTGTGCATTTCCCGGCATATTGGAACC	438		
DB	490	CAGGGAAATGCCGGCAGGCTGTGCCGCTTCGCCGCAATTCCAATGTGTTCAACC	549		

QY 439 GTTGAGTATAAGAAATTACTCTGCTGACCGAGGTATAGTTAAACACCTTCCTCCGT 494  
 DB 550 GTTATCATCGGCAGGCATACATATGTTGACGGCGTCTGTGCGACGCCGTGCGCGT 605

## RESULT 4

ID AA212213 standard; DNA; 903 BP.  
 AC AA212213;  
 XX 08-OCT-1999 (first entry)  
 XX Neisseria meningitidis complete ORF137 sequence.  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW Neisseria meningitidis; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.  
 XX Neisseria meningitidis.  
 OS Neisseria meningitidis.  
 XX WO9924578-A2.  
 XX 20-MAY-1999.  
 XX 09-OCT-1998; 98WO-IB01665.  
 XX 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX (CHIR-) CHIRON SPA.  
 XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 PI WPI; 1999-327407/27.  
 XX P-PSDB; AAY38778.

Proteins from Neisseria meningitidis and N. gonorrhoeae useful for diagnosis, treatment and prevention of infection  
 Claim 9; Page 323; 524pp; English.  
 XX Nucleotide sequences AA211972-212358 represent open reading frames (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their fragments, their nucleic acids and antibodies are used for diagnosis, prevention (as vaccines) or treatment of Neisseria infections, such as meningitis, septicaemia and gonorrhea. Both organisms are closely related. Fragments of the nucleic acids are useful as hybridisation probes and antisense reagents.  
 XX Sequence 903 BP; 204 A; 230 C; 255 G; 214 T; 0 other;  
 SQ Query Match 6.9%; Score 52; DB 20; Length 903;  
 Best Local Similarity 44.3%; Pred. No. 1.8e-06;  
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGATAAACTCGTTCCTTCGCGAGGAGCTGCAAGGCGATAGCCACATAGTGT 78  
 DB 130 GCAGTGTTCGGTTTGGCATCTCGTGGCGGCATCTAAGGATTTGCCATGTAGTATT 189  
 QY 79 TTGAAAGCTATAAACGAGCTCGGTATAGGGTGAGGGCTTTAAGCGGGGTAGCGCGG 138  
 DB 190 ATTAAGCTTTTGAAGAAAAACGGTATTCTCTGTGAAGGTGTTACCGGCACATCGCGAGGT 249  
 QY 139 GCATCTCTTCGTTCTTTATGCTCAGGCTACTCCCTGAAGGGATGTTACGCTTCG 198  
 DB 250 TCGATTGCGGCACCTTTTTCATCGGTATGTCGCCGACCGCTCGAANTGGNAGCC 309

QY 199 AAGAGGTAAACTGCTGAGCTGTTAAGTTCAAGCCACCTCTCAAGGGATTGATAGG 258  
 DB 310 GAAATTTTAGGCAAAACCGATTGTCGATTAAACCTTGTCCACAGTGGTTTATCAA 369  
 QY 259 TCGGGAAGGCTATAAGATTCTCTCAGGAAGTTCTCCCTTACAGGAGAAATAGAAAATT 318  
 DB 370 GCGGAAGCTCAAAATTAACATCAACCGAAAGTCGGCGGCGAGCATTCAGCAGTTT 429  
 QY 319 GAGATACGACATATATATGCGGACGGAATTATTAATCTGGGAGGGCTCTATACCTCTCG 378  
 DB 430 CCCATCAAAATTTGCGCGCGTTCCTACTGATTTTGAACCGGCAAGCGCGTTCCTTCAAT 489  
 QY 379 GAAGGAGTTTAATCCCGCACTTCTCGGCACTGTGCAATTCCTCGGCATATTTCGAACCC 438  
 DB 490 CAGGGGATCGCGGCGAGCTGTGCGGCTTCGCGGCATTCCTCAATGTGTTCCACCC 549  
 QY 439 GTTGAGTATAAGAAATTACTTGTCTGTGACGGAGTATAGTTAAACACCTTCCTCGT 494  
 DB 550 GTTATCATCGGCGGCATACATATGTTGACGGCGTCTGTGCGACGCCGTGCGCGT 605

## RESULT 5

AAA81473/c  
 ID AAA81473 standard; DNA; 92934 BP.  
 XX AAA81473;  
 AC AAA81473;  
 XX 04-DEC-2000 (first entry)  
 XX N. meningitidis partial DNA sequence gnm\_21 SEQ ID NO:21.  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX Neisseria meningitidis.  
 XX WO200022430-A2.  
 XX 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US23573.  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR) CHIRON CORP.  
 FA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.

Isolated nucleotide sequences of Neisseria meningitidis which can be used in the diagnosis and treatment of N. meningitidis infection and other Neisserial infections, for example, N.gonorrhoea -  
 Claim 7; Page 471-498; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AAA81453 to AAA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to



CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*: against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

SO Sequence 92934 BP; 24051 A; 24249 C; 21902 G; 22730 T; 2 other;

Query Match 6.9%; Score 52; DB 21; Length 92934;  
 Best Local Similarity 44.3%; Pred. No. 2.2e-05;  
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGATTAACCTGCTTTCTGGGAGAGCTGCAAGGCAATAGCCCATAGGTGTT 78  
 DB 50133 GCAATGCTGCTTTGCACTGCGGCGCATTAAGGATTTGCCATGTAGTATT 50074  
 QY 79 TTGAAGCTATTAACGAGCTCGATTAAGGGTGAAGGCTTTAAGCGGGTGAGCCGGG 138  
 DB 50073 ATTAAAGTTTGAAGAAAGGATTCCTGTGAAGGTGTTACCGGCACATCGGAGGT 50014  
 QY 139 GCAATGCTTTGCTTTTATGCTTCAAGGCTACCTCCCTTAAGAGGATTTGACCTTCG 198  
 DB 50013 TCGATTTTCGCGACCTTTTTCATCGGGTATGTCGCCACCGCTCGAATTGGAAGCC 49954  
 QY 199 AAGAGGGAACCTGCTGGAAGCTGTTAAGTTCAAGCCACCTCGAAGGATTTGATAGG 258  
 DB 49953 GAAATTTTAGGCAAAACCGATTTGCTGATTAACCTTTCACACGATTTTATCAAA 49894  
 QY 259 TGGAGAGAGCTATTAAGATTCCTTGAAGAAATTCCTTTTACGAGAAATTAAGAACTT 318  
 DB 49893 GGGCAAAAGCTGCAAAATTAATCATCAACGAAAGTGGCGGAGGATTTGACGATTT 49834  
 QY 319 GAGATTAACCGATTAATATGCGGAGGATTTATCTCGGGAAGGGCTCTATACCTCTCG 378  
 DB 49833 CCCATCAAAATTTGCCCGCTGCTACTATGATTTTGAACCGGCAAGGCGCTTCATAT 49774  
 QY 379 GAAGGAGTTTAATCCCGCACTTCGCGAGCTGTCGCAATTCGCCGATATTGAAGCC 438  
 DB 49773 CAGGGGAATGCGGCGAGGCTGTGCGGCTTCGCGCCCATTCCTCAATGTGTCAACCC 49714  
 QY 439 GTTGAATTAAGATTAATCTGCTGCTTGAAGAGGATTTAAGTTAACAATTCCTCGT 494  
 DB 49713 GTTATCAATCGGCAAGCATATATGTTGACGGCGGTCTGTGCGAGCCGTCGCGCT 49658

RESULT 6  
 AAF21613/c  
 ID AAF21613 standard; DNA; 172325 BP.

XX AAF21613;  
 DT 13-MAR-2001 (first entry)  
 DE *Neisseria meningitidis B* nucleotide sequence SEQ ID NO:114.  
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;  
 KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KM ds.  
 XX *Neisseria meningitidis*.  
 OS  
 XX MO20006791-A1.  
 PN  
 XX 09-NOV-2000.  
 PF 08-MAR-2000; 2000WO-US05928.

XX 30-APR-1999; 99US-0132068.  
 PR 08-OCT-1999; 99WO-US23573.  
 PR 28-FEB-2000; 2000GB-0004695.  
 PA (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 PI Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC, Masigiani V,  
 PI Galocci C, Mora M, Ratti G, Scarselli M, Scarlatto V, Rappuoli R,  
 PI Frazer CM, Grandi G;  
 DR WPI; 2000-647603/62.

PT *Neisseria meningitidis B* full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent *Neisserial* infections -  
 PS Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of  
 CC *Neisseria meningitidis B* (NMB). The sequences in AAF21544 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
 CC *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21589 to  
 CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to *Neisserial* bacteria or as a diagnostic reagent for detecting the  
 CC presence of *Neisserial* bacteria or of antibodies raised to *Neisserial*  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.

SO Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;

Query Match 6.9%; Score 52; DB 21; Length 172325;  
 Best Local Similarity 44.3%; Pred. No. 3e-05;  
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGATTAACCTGCTTTCTGGGAGAGCTGCAAGGCAATAGCCCATAGGTGTT 78  
 DB 153800 GCAATGCTGCTTTGCACTGCGGCGCATTAAGGATTTGCCATGTAGTATT 153741  
 QY 79 TTGAAGCTATTAACGAGCTCGTATTAAGGTGAGGCTTTAAGCGGGGTGAAGCCCGG 138  
 DB 153740 ATTAAAGTTTGAAGAAAGGATTCCTGTGAAGGTGTTACCGGCAATCGGACAGT 153681  
 QY 139 GCAATGCTTTGCTTTTATGCTTCAAGGCTACCTCCCTGAAGGATTTGACCTTCG 198  
 DB 153680 TCGATTTTCGCGACCTTTTTCATCGGGTATGTGCGCCGACCTCGAATTGGAAGCC 153621  
 QY 199 AAGAGGATTAACCTGCTGGAAGCTGTTAAGTTCAAGCACTTCGAGGATTTGAAGG 258  
 DB 153620 GAAATTTTAGGCAAAACCGATTTGCTGATTTAATCTTGTCCACAGTGTATTATCAA 153561  
 QY 259 TGGAGAGAGCTATTAAGATTCCTTGAAGAAATTCCTTACGAGGATTAAGAACTT 318  
 DB 153560 GCGGAAAGCTGCAAAATTAATCAACGAAAGTGGCGGAGGACAGATTCAGAGTTT 153501  
 QY 319 GAGATTAACGAGTATTAATATGCGGAGGATTTATCTCGGGAAGGGCTCTATACCTCG 378  
 DB 153500 CCATCAAAATTTGCCCGCTGCTACTATGATTTTGAACCGGAGAGGCGCTTCATAT 153441  
 QY 379 GAAGGAGTTTAATCCCGCACTTCCTGCGACGCTGTGCAATTCGCGCATATTGAAGCC 438



PT biodiversity  
 XX  
 PS Claim 1, SEQ ID No 13751, 103bp, English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1877 BP; 426 A; 464 C; 516 G; 470 T; 1 other;  
 Query Match 6.6%; Score 49.8; DB 23; Length 1877;  
 Best Local Similarity 52.7%; Pred. No. 1.4e-05;  
 Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
 QY 340 GCGACGATTAATCTCTCGGAAAGGCGCTTACCTTCGGAAGGAGTTTATCCCGCA 399  
 DB 54 GCCACCAATTAATAGGAGCGATGATTTGTTTCTGAAGGCATCTTCATCTTGCT 113  
 QY 400 CTCTCGGACGCTGTGCAATTCGCGCATATTGTAACCCGTTGAGTAAAGAAATTA 459  
 DB 114 ATTGCGCATCATCATGATATTCAGACTCATGCACTGTGCACTTAACGGCTACTGG 173  
 QY 460 CTGCTTACGAGGATTAAGTTAAACCTTCGCTTACGCTTTCAGGAAAGCGTAT 519  
 DB 174 CTGGTTATAGAGAGATGCTTAAACCAATTCCTTCTCCTCAGCGCTGATTTGGGGCT 233  
 QY 520 CCCACCGTTTCGTTGATGTCCTTC 544  
 DB 234 GATATTGTGATGACGCTTGACCTGC 258  
 RESULT 9  
 AA121215  
 ID AA121215 standard; DNA; 903 BP.  
 XX  
 XX AA121215;  
 AC  
 XX 08-OCT-1999 (first entry)  
 DT  
 XX  
 DE Neisseria gonorrhoeae complete ORF137 sequence.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.  
 OS Neisseria gonorrhoeae.  
 XX  
 PN MO9924578-A2.  
 XX  
 XX 20-MAY-1999.  
 PD  
 XX  
 PF 09-OCT-1998; 98WO-1B01665.  
 XX  
 PR 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PI  
 DR Grandi G, Maegnan V, Pizza M, Rappoli R, Scarlato V;  
 DR MPI, 1999-327407/27.  
 DR P-PSDB; AAY38780.  
 XX  
 PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 XX diagnosis, treatment and prevention of infection  
 PS Claim 9; Page 324-325; 524pp; English.  
 CC Nucleotide sequences AA211972-21358 represent open reading frames  
 CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of Neisseria infections,  
 CC such as meningitis, septicemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 903 BP; 207 A; 228 C; 257 G; 211 T; 0 other;  
 Query Match 6.3%; Score 47.4; DB 20; Length 903;  
 Best Local Similarity 43.6%; Pred. No. 5.9e-05;  
 Matches 213; Conservative 0; Mismatches 276; Indels 0; Gaps 0;  
 QY 19 GAAGAGATAAACTGCTCTTCCGAGAGAGCTGCAAGGCAATAGCCACATAGTGT 78  
 DB 130 GCAAGTGTGCTGCTTGGCACTGGTGGCGGCATTAAGATTTGCCATATAGAT 189  
 QY 79 TTGAAGCTATTAACGAGCTCGTATTAAGGCTGAAGGCTTTAAGCGGGGTAGCGCCGG 138  
 DB 190 GTTAAGGTTTGAAGAAAGAAACGGTATCTGTGAAGGTGTATACCGGACATCGGACAGT 249  
 QY 139 GCAATCGTTCGCTTTTATGCTCAGGCTACCTCCCTGAAGGATTTACGCTTCTG 198  
 DB 250 TCGATATGTCGAGCGCTTTTGGCATCGGTATGTCGCCGACCGCTGAAATGGAAGCC 309  
 QY 199 AAGAGGTAATGCTGCTAGAGCTGTTAAGTTCAAGCCACTCTGAAGGATATAGGG 258  
 DB 310 GAGATTTTATGTAACCGATTTAGTCATTTAATCTTGTCCACAGTGTATATCAA 369  
 QY 259 TGGAGAAAGCTATTAAGATTCCTTGAGAGAGTTCCTTACAGAGATAGAAAACCT 318  
 DB 370 GCGGAAAGCTGCAAAATTAATCAACGAAAGTCGCGGAGGCAATTCAGAGATT 429  
 QY 319 GAGATACGACGTATATATGCGGACGATTTATCTGGAAAGGCTCTATACCTTCG 378  
 DB 430 CCCATCAATTTCCGCGCGTTCACATATTTTGAACCGGCAAGCCGCTGCTTCAAT 489  
 QY 379 GAAGGAGTTTAATCCCGCACTTCTCGGAGCTGTGCAATTCGCGCATATTTGAACCC 438  
 DB 490 CAAAGGAATGCGGGAGAGCGGTGTGCTTCCGCGCATTTCCAAATGTTCAGGCA 549  
 QY 439 GTTGAATTAAGATTAATCTGCTGTTACGAGGATTAAGTTAACAACCTTCGTTAG 498  
 DB 550 GTTCATCATCGGAGGACAAATATGTTTGAACGCGGCTGTGTGCAACCCGTCGCTCAGT 609  
 QY 499 CCCTTCAG 507  
 DB 610 GCCGCTCG 618  
 RESULT 10  
 AAV81946  
 ID AAV81946 standard; DNA; 40138 BP.  
 XX

AAV81946;  
 19-OCT-1999 (first entry)  
 V. marinus PKS-like cluster comprising ORFs 6, 7, 8 and 9.  
 Polyketide-like synthesis; PKS; PKS-like gene; PUPA; DHA; transgenic;  
 poly-unsaturated fatty acid; eicosapentenoic acid; docosahexenoic acid;  
 EPA; oil; dietary supplement; infant feeding formulation; malnutrition;  
 intravenous feeding formulation; cooking oil; fat; anti-inflammatory;  
 cholesterol; open reading frame; ORF; ss.  
 Vibrio marinus.

Key Location/Qualifiers  
 CDS 17394..25352  
 /tag= a  
 /product= "ORF 6"  
 CDS 25509..28160  
 /tag= b  
 /product= "ORF 7"  
 CDS 28209..34265  
 /tag= c  
 /product= "ORF 8"  
 CDS 34454..36118  
 /tag= d  
 /product= "ORF 9"

WO985625-A1.  
 10-DEC-1998.  
 04-JUN-1998; 98WO-US11639.  
 04-JUN-1997; 97US-0048650.  
 (CALJ ) CALGENE LLC.  
 Facciotti D, Lassner M, Metz JG;  
 WPI; 1999-070271/06.  
 New nucleic acid encoding polyketide-like synthesis enzymes of  
 Vibrio marinus and transformed plants and microbes that produce  
 polyunsaturated fatty acids, useful as pharmaceuticals and food  
 supplements  
 Example 1; Fig 5; 153pp; English.

The invention provides polyketide-like synthesis (PKS)-like genes that  
 are used for the production of long chain polyunsaturated fatty acid  
 (PUFA) productions. Genes responsible for eicosapentenoic acid (EPA)  
 production in Shewanella putrefaciens and novel genes associated with the  
 production of docosahexenoic acid (DHA) in Vibrio marinus are used to  
 generate transgenic plants that can express transgenes encoding PKS-like  
 genes associated with PUFA production. The PKS-like genes are used to  
 transform plants and microbial cells to give recombinants having altered  
 contents of PUFA (specifically DHA and EPA). Oils from these plants are  
 useful as dietary supplements (in infant feeding formulations, to give a  
 PUFA profile closer to that of human milk; for treating malnutrition; in  
 intravenous feeding formulations; in cooking oils, fats etc.), also as  
 anti-inflammatory agents and for reducing cholesterol levels. Fragments  
 from the genes are useful as probes to isolate related molecules. or to  
 detect organisms that express PKS-like genes. The method facilitates  
 large scale production of PUFA by providing new pathways for their  
 synthesis or suppressing interfering pathways. Expression of PUFA in  
 seeds allows simple recovery, as oil which can be engineered to have a  
 particular PUFA profile. Expression in microbes also allows simple  
 recovery and control of PUFA profile and is not subject to external  
 variables such as weather or food supply. The present sequence represents  
 an approximately 40 kb PKS-like cluster DNA fragment from V. marinus  
 comprising open reading frames (ORFs) 6, 7, 8 and 9.

SQ Sequence 40138 BP; 11845 A; 7635 C; 8758 G; 11900 T; 0 other;  
 Query Match 6.1%; Score 46.2; DB 20; Length 40138;  
 Best Local Similarity 52.9%; Pred. No. 0.0012;  
 Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
 QY 23 AGATAAACCTCGTTCTTTTCGGGAGGAGCTGCACAAAGGCATAGCCACATAGGTGTTCGA 82  
 Db 14817 AAATTGGCTTAGTCTTCTTCTGCGGTGGTGGCAAGGATTTGCTCATCTTGGTGTATTAA 14876  
 QY 83 AAGCTATAACAGAGCTCGGTATAGCGGTGAGGGCTTTTAAAGCGGGGTGAGCGCGGGGCAA 142  
 Db 14877 AATACCTGTAGAGCAAGATATAAGACCGAATGTAAATTCGGGTACAAAGTGTGCTCTTA 14936  
 QY 143 TCGTTTCGTCCTTTATGCTCAGGCTACTCCCTGAAGGATGTTTCAGCCTTCTGAAGA 202  
 Db 14937 TCGTTGGTGCACTTTTATGCTCAGGACTTGAGATTGATGACATTTTACAAATTCCTTCG 14996  
 QY 203 GGGTAAA 209  
 Db 14997 ATGTAAA 15003  
 RESULT 11  
 AAA711520  
 ID AAA71520 standard; DNA; 40138 BP.  
 AC AAA71520;  
 XX  
 DT 11-DEC-2000 (first entry)  
 DE V. marinus PKS-like gene cluster encoding ORF6 to ORF9.  
 KW PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;  
 KW polyketide-like synthesis; PUFA; dietary supplement; intravenous feeding;  
 KW malnutrition; cooking oil; cooking fat; margarine;  
 KW docosahexenoic acid production; eicosapentenoic acid production; ds.  
 XX  
 OS Vibrio marinus.  
 XX  
 PN WO200042195-A2.  
 XX  
 PD 20-JUL-2000.  
 XX  
 PF 14-JAN-2000; 2000WO-US00956.  
 XX  
 PR 14-JAN-1999; 99US-0231899.  
 XX  
 PA (CALJ ) CALGENE LLC.  
 XX  
 PI Facciotti D, Metz JG, Lassner M;  
 XX  
 DR WPI; 2000-476063/41.  
 XX  
 PT New DNA sequences encoding for polyketide (PK)-like synthesis pathway  
 PT genes from Shewanella, Vibrio and Schizochtrium, useful for creating  
 PT transgenic plants that express poly-unsaturated long chain fatty acids  
 XX  
 PS Example 1; Fig 5; 302pp; English.  
 XX  
 CC This invention describes novel DNA sequences encoding for polyketide  
 CC (PK)-like synthesis (PKS-like) pathway genes from Shewanella, Vibrio  
 CC and Schizochtrium. The nucleic acids are useful for isolating related  
 CC molecules or in methods to detect organisms expressing the PKS-like  
 CC genes. They are also useful for creating transgenic plants that express  
 CC poly-unsaturated long chain fatty acids. The poly-unsaturated long chain  
 CC fatty acids produced recombinantly are useful as dietary supplements for  
 CC patients undergoing intravenous feeding or for preventing or treating  
 CC malnutrition. The poly-unsaturated long chain fatty acids can also be  
 CC incorporated into cooking oils, fats or margarine formulated so that in  
 CC normal use the recipient receives a desired amount of poly-unsaturated  
 CC long chain fatty acids. The nucleic acids are also useful in large scale

CC production of docosahexaenoic acid and eicosapentaenoic acid, and for the  
CC modification of the fatty acid profile of host cells and edible plant  
CC tissues and/or plant parts. Transgenic production of polyunsaturated  
CC fatty acids in particular host cells allows quicker purification from  
CC natural sources such as fish or plants. This sequence represents a  
CC *Vibrio* marinus DNA fragment encoding the PKS gene cluster incorporating  
CC ORF6 to ORF9.

XX  
CC  
SQ Sequence 40138 BP; 11845 A; 7634 C; 8759 G; 11900 T; 0 other;

Query Match 6.1%; Score 46.2; DB 21; Length 40138;  
Best Local Similarity 52.9%; Pred. No. 0.0012;  
Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 23 AGATTAACCTGCTTCTTGGGAGGAGCTGCAAGGCGCATAGCCCATAGGCTTTGA 82

DB 14817 AAATGGCTTAGCTCTTCTTGGGAGGAGCTGCAAGGCGCATAGGCTTTGA 14876

QY 83 AAGCTATTAACGAGCTGCTATTAAGGCTTAAAGCGGGTGAGCGCCGGGCA 142

DB 14877 AATACCTTTAGAGCAAGATTAAGACCGATTAATGCGGCTACAGGCTGCTTA 14936

QY 143 TCGTTTCGCTTTTATGCTTCAAGCTACCTCCCTGAAGGAGTTCAGGCTTGAAGA 202

DB 14937 TGGTTGGTCACTTATTTGCTCAGACTTGAGATTATGATTTACATTTCTTCATCG 14996

QY 203 GGGTAA 209

DB 14997 ATGTAA 15003

RESULT 12

AA164984  
ID AA164984 standard; DNA; 41587 BP.

XX  
AC AA164984;

XX  
DT 04-DEC-2001 (first entry)

XX  
DE Moritella marina icosapentaenoic acid biosynthesis enzyme DNA #1.

XX  
KW Icosapentaenoic acid biosynthesis; docosahexaenoic acid productivity;

XX  
KM DNA industrial production; ds.

XX  
OS Moritella marina.

XX  
PN JP2001169780-A.

XX  
PD 26-JUN-2001.

XX  
PF 15-DEC-1999; 99JP-0356614.

XX  
PR 15-DEC-1999; 99JP-0356614.

XX  
PA (KEIZ-) KEIZAI SANGYOSHO SANGYO GIJUTSU SOGO KEN.

XX  
XX  
XX WPI; 2001-592536/67.

XX  
PT Gene of a docosahexaenoic acid-producing microbe, used for industrial

XX  
PT production of docosahexaenoic acid -

XX  
PS Claim 4; Page 10-24; 68pp; Japanese.

XX  
CC The present invention provides a DNA sequence encoding an icosapentaenoic

CC acid-biosynthesizing enzyme group-like protein group derived from a

CC microbe having docosahexaenoic acid (DHA) productivity. The gene can be

CC used for the industrial production of DHA. The present sequence is a

CC coding sequence described in the exemplification of the invention.

Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 23 AGATTAACCTGCTTCTTGGGAGGAGCTGCAAGGCGCATAGCCCATAGGCTTTGA 82

DB 7433 AAATGGCTTAGCTCTTCTTGGGAGGAGCTGCAAGGCGCATAGGCTTTGA 7492

QY 83 AAGCTATTAACGAGCTGCTATTAAGGCTTAAAGCGGGTGAGCGCCGGGCA 142

DB 7493 AATACCTTTAGAGCAAGATTAAGACCGATTAATTCGGGCTACAGGCTGCTTA 7552

QY 143 TCGTTTCGCTTTTATGCTTCAAGCTACCTCCCTGAAGGAGTTCAGGCTTGAAGA 202

DB 7553 TGGTTGGTCACTTATTTGCTCAGACTTGAGATTATGATTTACATTTCTTCATCG 7612

QY 203 GGGTAA 209

DB 7613 ATGTAA 7619

RESULT 13

ABK79457  
ID ABK79457 standard; DNA; 467 BP.

XX  
AC ABK79457;

XX  
DT 13-AUG-2002 (first entry)

XX  
DE Bacillus clausii genomic sequence tag (GST) #2300.

XX  
KW Differential gene expression; genomic sequenced tag; GST;

XX  
KW altered culture condition; environmental stress;

XX  
KW physiological provocation; ds.

XX  
OS Bacillus clausii.

XX  
PN WO200229113-A2.

XX  
PD 11-APR-2002.

XX  
PF 05-OCT-2001; 2001WO-US31437.

XX  
PR 06-OCT-2000; 2000US-0680598.

XX  
PR 27-MAR-2001; 2001US-279526P.

XX  
PA (NOVO) NOVOZYMES BIOTECH INC.

XX  
PA (NOVO) NOVOZYMES AS.

XX  
PI Berka R. Clausen IG;

XX  
XX  
XX WPI; 2002-416684/44.

XX  
DR Claim 11; SEQ ID NO 6748; 200pp; English.

XX  
CC The invention describes a method of monitoring differential expression of

CC genes in a first Bacillus cell relative to expression of the genes in

CC other Bacillus cells, comprising hybridising labelled nucleic acid probes

CC isolated from Bacillus cells to a substrate containing array of Bacillus

CC genomic sequenced tags (GST), examining the array, and determining

CC relative gene expression by an observed hybridisation reporter signal of

CC a spot in the array. The method is useful for measuring the expression of

CC genes in a first Bacillus cell relative to expression of the same genes

CC in one or more second Bacillus cells. The method is useful for monitoring

CC global expression of several genes from a Bacillus cell, discovering new

CC genes, identifying possible functions of unknown open reading frames and

CC monitoring gene copy number variation and stability. Monitoring changes

CC in expression of genes may be used to provide a representation of the way

CC in which Bacillus cells adapt to changes in culture conditions.

CC environmental stress or other physiological provocation. Extensive

CC follow-up characterisation is unnecessary, when one spot on an array  
CC equals one gene or one open reading frame, since sequence information is  
CC available. This sequence represents a genomic sequence tag (GST) used in  
CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 467 BP; 122 A; 88 C; 133 G; 124 T; 0 other;

Query Match 6.1%; Score 45.8; DB 24; Length 467;

Best Local Similarity 44.4%; Pred. No. 0.00014;

Matches 185; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 24 GATTAACCTCGTCTTTTCGGAGAGAGCTGCAAGGGCATAGCCACATAGGTGTTTGA 83

DB 21 GATTGGACTGGCCCTTGGGCTCTGGCGGGCTAGAGGCTTTGCCACATTTGGTGTAA 80

QY 84 AGCTATAACAGAGCTCGGTATAAGGGTGAGGGCTTTAAGCGGGGTGAGCGCGGCAAT 143

DB 81 GACGCTTACAGAGCGGGGTTACATGTACTATTAGCTGGTAGCAGCATGGTGCAAT 140

QY 144 CGTTTCGCTCTTTATGCTCAGGCTACTCCCTGAAGGATGTTTCAGCCCTTCTGAAGAG 203

DB 141 GGTGCGGACTATGATGCGTAGGCCATTGCGTTGAAACGATGGAAGTTTTCGGAACA 200

QY 204 GGTAACTGGCTGAAGCTGTTTAAGTTCAAGCCACCTCTGAAGGATGTAGGGTGGGA 263

DB 201 TTTCAAAACGAAATATTTACTTTGGATTTTACAGTCTCAAAACAAAGGCTTATGCTGACA 260

QY 264 GAAGGCTATAAGATTCTCTGAGGAAGTTCTCCCTTACAGGAGATAGAAAACTTGAGAT 323

DB 261 AAGATTGAAGGCTCATTCGTTTACTAGCTTAAAGACAAAGCTGGAGTCGCTTTTTC 320

QY 324 ACCGACGATATATGCGGAGGATTTATATCTCGGGAAGGCTCTATACCTCTCGGAAG 383

DB 321 GCCAGTTCAAGTGTAGCACTGATCTTTAAGTGGCCAAAAGTGTGTATGTCAAAGG 380

QY 384 GAGTTTAATCCCGCACTTCTCGGACGTGTGCAATTTCCCGGCATATTTGCAACCCGT 440

DB 381 AGACGTCGCCAAGCGGTTGAGCAAGCTTGTGATTCAGGCAATTTTGTCTCTGT 437

RESULT 14

AAZ12212

ID AAZ12212 standard; DNA; 447 BP.

XX AC AAZ12212;

XX DT 08-OCT-1999 (first entry)

XX DE Neisseria meningitidis partial ORF137 sequence.

XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.

XX OS Neisseria meningitidis.

XX XX WO9924578-A2.

XX XX 20-MAY-1999.

XX XX 09-OCT-1998; 98WO-IB01665.

XX XX 01-SEP-1998; 98GB-0019016.

XX PR 06-NOV-1997; 97GB-0023516.

XX PR 14-NOV-1997; 97GB-0024190.

XX PR 18-NOV-1997; 97GB-0024386.

XX PR 27-NOV-1997; 97GB-0025158.

XX PR 10-DEC-1997; 97GB-0026147.

XX PR 14-JAN-1998; 98GB-0000759.

XX XX

PA (CHIR-) CHIRON SPA.

XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;

XX WPI; 1999-327407/27.

DR P-PSDB; AAY38777.

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for

PT diagnosis, treatment and prevention of infection

XX Claim 9; Page 322; 524pp; English.

XX Nucleotide sequences AAZ11972-Z12358 represent open reading frames

CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode

CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their

CC fragments, their nucleic acids and antibodies are used for diagnosis,

CC prevention (as vaccines) or treatment of Neisseria infections,

CC such as meningitis, septicaemia and gonorrhea. Both organisms

CC are closely related. Fragments of the nucleic acids are useful

CC as hybridisation probes and antisense reagents.

XX SQ Sequence 447 BP; 110 A; 115 C; 114 G; 107 T; 1 other;

Query Match 5.6%; Score 42.6; DB 20; Length 447;

Best Local Similarity 54.0%; Pred. No. 0.0016;

Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 19 GAAGAGATAAACCTCGTCTTTTCGGAGGAGCTGCAAGGGCATAGCCACATAGTGT 78

DB 130 GCAGTGGTGGTTCGGCACTCGGTGGCGGCGCATCTAAAGGATTTGCCCATGTAGTATT 189

QY 79 TTGAAGCTATAAACGAGCTGGTATAAGGGTGAGGGCTTTAAGCGGGTGAGCCCGGG 138

DB 190 ATTAAGGTTTGAAGAAGAACGGTATTCCTGTAAGGTGGTTACCGGCACTCCCGAGGT 249

QY 139 GCAATCGTTTCGTCCTTTTATGCTCAGCTACTCCCTGA 179

DB 250 TCGATTGTCGCAACCTTTTTCATCGGGTATGTCGCCCGA 290

RESULT 15

AAA81390

ID AAA81390 standard; DNA; 447 BP.

XX AC AAA81390;

XX DT 04-DEC-2000 (first entry)

XX DE N. meningitidis MenB polynucleotide sequence ORF number 66.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

XX KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX KW Meningococcus B; MenB; ds.

XX OS Neisseria meningitidis.

XX XX WO200022430-A2.

XX XX 20-APR-2000.

XX XX 08-OCT-1999; 99WO-US23573.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

XX PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

XX PI Rappuoli R, Pizza M;

XX DR WPI; 2000-318079/27.

XX XX

PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
PT used in the diagnosis and treatment of *N. meningitidis* infection and  
PT other *Neisseria* infections, for example, *N. gonorrhoea* -

PS Disclosure; Page 216; 1760pp; English.

XX  
CC The present invention describes methods of obtaining immunogenic  
CC proteins from *Neisseria* genomic sequences. AA81453 to AA82414  
CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent  
CC *Neisseria* DNA sequences and their corresponding proteins; AA81254 to  
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the  
CC isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to  
CC AA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC *Neisseria* bacteria. For example, some of the identified proteins could  
CC be components of vaccines against *Meningococcus B*; against all serotypes;  
CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX  
SQ Sequence 447 BP; 110 A; 115 C; 114 G; 107 T; 1 other;

Query Match 5.6%; Score 42.6; DB 21; Length 447;

Best Local Similarity 54.0%; Pred. No. 0.0016;

Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 19 GAAGAGATAAACCTCGTTCTTTCGGGAGAGCTGCAAGGACATAGCCCATAGGTGTT 78  
DB 130 GAGAGTGTGCGTTTGGCAGCTCGTGGCGGCGCATCTAAGAGATTGGCCATGTAGGATTT 189  
QY 79 TTGAAGCTATTAACGAGCTCGGTATTAAGGCTGAGGCTTTAAGCGGAGTGAGCGCCGG 138  
DB 190 ATTAAGTTTGAAGAAACGATATTCCTGTGAAGGTGTTACCGGACACTCCGACGGT 249  
QY 139 GCAATCGTTTGGCTTTTATGACCTCAGGCTACTCCCTGA 179  
DB 250 TCGATTGCGGACACCTTTTTCATCGGGGTATGTGCCCA 290

Search completed: June 19, 2003, 13:38:25

Job time : 182 secs

Thu Jun 19 17:30:36 2003

us-09-903-410-26.rml

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 13:34:10 ; Search time 47 Seconds

(without alignments)  
4932.928 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756  
1 ttgagattcgagaatttga.....tagagttacttctgaaatag 756

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/5C\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/5D\_COMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	756	100.0	756	2	US-08-602-359A-26
2	397.8	52.6	750	2	US-08-602-359A-29
3	93.4	12.4	2108	4	US-09-221-017B-305
4	46.2	6.1	40138	3	US-09-090-793-12
5	40.4	5.3	7218	1	US-08-232-463-14
6	32.2	4.3	4403765	4	US-09-103-840A-2
7	32.2	4.3	4411529	4	US-09-103-840A-1
8	31.4	4.2	2057	3	US-09-008-303-1
9	31	4.1	2409	4	US-09-484-970B-101
10	30.4	4.0	1555	3	US-08-747-574-3
11	29.8	3.9	1867	2	US-08-607-509-3
12	29.8	3.9	1867	2	US-08-634-642-3
13	29.6	3.9	1867	2	US-08-989-370-3
14	29.6	3.9	756	2	US-08-602-359A-26
15	29.6	3.9	6735	4	US-08-561-527-104
16	29	3.8	14602	1	US-08-597-236-1
17	29	3.8	14602	1	US-08-746-682A-1
18	28.8	3.8	10763	1	US-08-761-258-1
19	28.8	3.8	10763	1	US-08-977-306-1
20	28.8	3.8	4403765	4	US-09-103-840A-2
21	28.8	3.8	4411529	4	US-09-103-840A-1
22	28.6	3.8	1803	4	US-08-976-259-71
23	28.6	3.8	5577	1	US-08-326-117B-1
24	28.6	3.8	5577	1	US-08-982-129-1
25	28.6	3.8	5582	4	US-09-178-176B-1
26	28.6	3.8	5582	4	US-09-457-864-1
27	28.4	3.8	510	4	US-09-328-111-722

C 28	28.4	3.8	1344	1	US-07-940-852-1	Sequence 1, Appli
C 29	28.4	3.8	1344	1	US-08-396-126-1	Sequence 1, Appli
C 30	28.4	3.8	1710	3	US-09-000-630C-1	Sequence 1, Appli
C 31	28.4	3.8	1710	3	US-08-862-730C-1	Sequence 1, Appli
C 32	28.2	3.7	951	1	US-08-671-525B-1	Sequence 1, Appli
C 33	28.2	3.7	951	1	US-08-672-109B-1	Sequence 1, Appli
C 34	28.2	3.7	951	1	US-08-842-045-1	Sequence 1, Appli
C 35	28.2	3.7	951	2	US-08-842-238-1	Sequence 1, Appli
C 36	28.2	3.7	951	3	US-08-629-335B-1	Sequence 1, Appli
C 37	28.2	3.7	1270	4	US-08-387-805-1	Sequence 1, Appli
C 38	28.2	3.7	1633	1	US-07-866-979-5	Sequence 5, Appli
C 39	28.2	3.7	1633	2	US-08-466-906B-5	Sequence 5, Appli
C 40	28.2	3.7	1633	3	US-08-706-281A-5	Sequence 5, Appli
C 41	28.2	3.7	1633	4	US-09-201-746-5	Sequence 5, Appli
C 42	28.2	3.7	1633	4	US-09-097-231-5	Sequence 5, Appli
C 43	28.2	3.7	1839	1	US-08-075-193-1	Sequence 1, Appli
C 44	28.2	3.7	1839	2	US-08-564-090A-1	Sequence 1, Appli
C 45	28.2	3.7	1839	5	PCT-US94-06698-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-602-359A-26  
Sequence 26, Application US/08602359A  
Patent No. 5942430  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MARFIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & RICHARDSON P. C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,359A  
FILING DATE: February 16, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 26:  
LENGTH: 756 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
US-08-602-359A-26



Query Match 100.0%; Score 756; DB 2; Length 756;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-241;  
 Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGAGTTCAGGAAATTTGAGAGATAAAGCTCGTTCTTCGGAGAGCTGCAAGGCG 60  
 Db 1 TTGAGTTCAGGAAATTTGAGAGATAAAGCTCGTTCTTCGGAGAGCTGCAAGGCG 60

Qy 61 ATAGCCCAATAGGTTGTTGAAAGCTATAAAGAGCTCGGTATAGGGTGAAGGCTTTA 120  
 Db 61 ATAGCCCAATAGGTTGTTGAAAGCTATAAAGAGCTCGGTATAGGGTGAAGGCTTTA 120

Qy 121 AGCGGGTGAGCGCGGGGCAATCGTTTCGGTCTTTATGCTTCAGGCTACTCCCTGAA 180  
 Db 121 AGCGGGTGAGCGCGGGGCAATCGTTTCGGTCTTTATGCTTCAGGCTACTCCCTGAA 180

Qy 181 GGGATGTTACGCTTCTGAAGAGGGTAAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACT 240  
 Db 181 GGGATGTTACGCTTCTGAAGAGGGTAAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACT 240

Qy 241 CTGAGGATGTTAGGTTGAGAGGCTATAAGATTCCTTGAGGAGTTCCTGCTTAC 300  
 Db 241 CTGAGGATGTTAGGTTGAGAGGCTATAAGATTCCTTGAGGAGTTCCTGCTTAC 300

Qy 301 AGGAGATAGAAAATTTGAGATACCGACGTATATATGCGCGACGATTTATCTCGGA 360  
 Db 301 AGGAGATAGAAAATTTGAGATACCGACGTATATATGCGCGACGATTTATCTCGGA 360

Qy 361 AGGGCTCTATACCTCTCGGAGGAGTTTAACTCCCGCACTTCGCGAGCTGTCGCAAT 420  
 Db 361 AGGGCTCTATACCTCTCGGAGGAGTTTAACTCCCGCACTTCGCGAGCTGTCGCAAT 420

Qy 421 CCGGCAATTTGAAACCGTTGAGTATAAGAAATTTACTTGTCTGCTGACGAGGTATAGTT 480  
 Db 421 CCGGCAATTTGAAACCGTTGAGTATAAGAAATTTACTTGTCTGCTGACGAGGTATAGTT 480

Qy 481 AACACCTTCGCTGAGGAGGTTTAACTCCCGCACTTCGCGAGCTGTCGCAAT 540  
 Db 481 AACACCTTCGCTGAGGAGGTTTAACTCCCGCACTTCGCGAGCTGTCGCAAT 540

Qy 541 CTTCCATAGAGCGGAAAGGATATAAAGAACTTCTTCAATCTCTTTGAGGAGCTTC 600  
 Db 541 CTTCCATAGAGCGGAAAGGATATAAAGAACTTCTTCAATCTCTTTGAGGAGCTTC 600

Qy 601 TTTCTTGGCTCGCTCAAACTCCGAAAGAGAGGAGTTTGTGACCTCGTTATAGTT 660  
 Db 601 TTTCTTGGCTCGCTCAAACTCCGAAAGAGAGGAGTTTGTGACCTCGTTATAGTT 660

Qy 661 CTTGAGCTTGAAGGAGTTCAACCCCTTGAATGTTAGAAAGCGGACCAAAATATGAGAGG 720  
 Db 661 CTTGAGCTTGAAGGAGTTCAACCCCTTGAATGTTAGAAAGCGGACCAAAATATGAGAGG 720

Qy 721 GGATACATAAAGGCTTAGAGTACTTTCTGAAATAG 756  
 Db 721 GGATACATAAAGGCTTAGAGTACTTTCTGAAATAG 756

RESULT 2  
 US-08-602-359A-29  
 ; Sequence 29, Application US/08602359A  
 ; Patent No. 5942430  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROBERTSON, Daniel E.  
 ; APPLICANT: MORPHY, Dennis  
 ; APPLICANT: REID, John  
 ; APPLICANT: MAFFIA, Anthony  
 ; APPLICANT: LINK, Steven  
 ; APPLICANT: SWANSON, Ronald V.  
 ; APPLICANT: WARREN, Patrick V.  
 ; APPLICANT: KOSMOTKA, Anna  
 ; TITLE OF INVENTION: ESTERASES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & RICHARDSON P.C.  
 STREET: 4225 EXECUTIVE SQUARE, STE 1400  
 CITY: LA JOLLA  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH DISKETTE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WORD PERFECT 6.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/602,359A  
 FILING DATE: February 16, 1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: HAILE, LISA A.  
 REGISTRATION NUMBER: 38,347  
 REFERENCE/DOCKET NUMBER: 09010/010001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-678-5070  
 TELEFAX: 619-678-5099  
 INFORMATION FOR SEQ ID NO: 29:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 750 NUCLEOTIDES  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: SINGLE  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE: GENOMIC DNA  
 US-08-602-359A-29

Query Match 52.6%; Score 397.8; DB 2; Length 750;  
 Best Local Similarity 72.4%; Pred. No. 3.2e-122;  
 Matches 516; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 1 TTGAGTTCAGGAAATTTGAGAGATAAAGCTCGTTCTTCGGAGAGCTGCAAGGCG 60  
 Db 13 TTGAAATTCAGAGAGGTTTGAAGAGGTTAACTTAGTTCTTCGGAGGCGCTGCCAAGGT 72

Qy 61 ATAGCCCAATAGGTTGTTGAAAGCTATAAAGAGCTCGGTATAAGGGTGAAGGCTTTA 120  
 Db 73 ATGCGCCATATAGTGTGTTTAAAGCTCTGAGAGCTCGGTATAAGGTTAAAGAGCTC 132

Qy 121 AGCGGGTGAGCGCGGGGCAATCGTTTCGGTCTTTTATGCTTCAGGCTACTCCCTGAA 180  
 Db 133 AGCGGGTAAAGTGTGAGAGCTATCGTTTTCGGTCTTTTACGCTTCGGGCTACATCCCGAC 192

Qy 181 GGGATGTTCAAGCTTCGTAAGAGGCTAAACTGGCTGAAAGCTGTTTAAAGTTCAAGCCACT 240  
 Db 193 GAGATGTTAAACTCTGAAAGAGCTAAACTGGCTCAAACTTTTAAAGTTCAAAACACCG 252

Qy 241 CTGAGGAGTATAGGTTGAGAGGCTATAAGATTCCTTGAGGAGTTCTCCCTTAC 300  
 Db 253 AAAATGGCTTAAATGGGGTGGGAGAGGCTGACAGGTTTGGAAAAAGAGCTCGGAGTT 312

Qy 301 AGGAGAAATAGAAAACCTTCAGATACCGAGCTATATATATCGCGAGGAGTTTATCTCGGA 360  
 Db 313 AAGAGGCTGGAAGACCTGAACATACCAACCTATCTTTCTCGCGGATCTGTACACGGA 372

Qy 361 AGGGCTTATACCTCTCGGAGGAGTTTAACTCCCGCACTTCCTCGGAGCTGTCGCAAT 420  
 Db 373 AAGGCTCTTACTTCGGAGAGGTTGACTTAAATTCCTCGGCTTCCTCGGAGGTTGTTCCATA 432

Qy 421 CCGGCAATTTGAAACCGTTGAGTATAAGAAATTTACTTGTCTGTTGAGCGAGGTATAGTT 480  
 Db 433 CCGGCAATTTTGAACCAAGTTCAGTACGAGAAATTTTCTACTTTTGAACGAGGTATAGTG 492

Qy 481 AACACCTTCGCTGAGGAGGCTTCAGGAAAGCGGTATTCGACCGTTTGGTGTGATGTC 540  
 Db 493 AACACCTTCGCTGAGGAGGCTTCAGGAAAGCGGTATTCGACCGTTTGGTGTGATGTC 552

Thu Jun 19 17:30:36 2003

us-09-903-410-26.rml

Page 3

QY 541 CTTCCATGAGCGGAGAAAGATATTAAGAAATTTTCACTCTTTTGGAGAGCTTC 600  
Db 553 CTTCCATGAGCGGAGAAAGATATTAAGAAATTTTCACTCTTTTGGAGAGCTTC 612  
QY 601 TTTCTGCGGCTGCTCAAACTCCGAAAGAGAGTTTGTGACCTGTTATAGTT 660  
Db 613 TTTCTGCGGCTGCTCAAACTCCGAAAGAGAGTTTGTGACCTGTTATAGTT 672  
QY 661 CTTGAGTTGAGAGTTCAACCCCTTGATGTTAGAAAAGCGAACCAATAT 713  
Db 673 CTTGAGTTGAGAGTTCAACCCCTTGATGTTAGAAAAGCGAACCAATAT 725

RESULT 3

US-09-221-017B-305  
Sequence 305, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Rose, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FORBSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
Prior Application DATA:  
APPLICATION NUMBER: P01182  
FILING DATE: 31-DEC-1997  
Prior Application DATA:  
APPLICATION NUMBER: P01546  
FILING DATE: 30-JAN-1998  
Prior Application DATA:  
APPLICATION NUMBER: P02911  
FILING DATE: 09-APR-1998  
Prior Application DATA:  
APPLICATION NUMBER: P07/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 305:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2108 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1...2108  
US-09-221-017B-305

Query Match 12.4%; Score 93.4; DB 4; Length 2108;  
Best Local Similarity 46.5%; Pred. No. 6,66-21;  
Matches 338; Conservative 0; Mismatches 386; Indels 3; Gaps 1;

QY 24 GATTAACCTGTTCTTTGAGGAGAGCTGCAAAAGGAGATATGACCAATAGGTTTGA 83  
Db 1247 GATTAACCTGTTCTTTGAGGAGAGCTGCAAAAGGAGATATGACCAATAGGTTTGA 1306  
QY 84 AGCTATTAACGAGCTGCTGATTAAGGAGGCTTTAAAGCGGAGTGAAGCGGAGAT 143  
Db 1307 TGCTCTTGAAGCAAAAGGAGATGCTCCGACATCTATGACAGTACAGTCCAGAGGCTT 1366  
QY 144 CGTTGGCTCTTTATGCTCAGGCTACCTCCCTGAAGGAGATGTTGACGCTTGAAAG 203  
Db 1367 GGTAGCTGACCTTATGCGGAGATGATACAGGCGGAGAGATGTTGACGCTTGAAAG 1426  
QY 204 GGTAACTGAGCTGAGAGCTGTTAAGTTCAAGCACTTGTGAAGGATTTGATAGGAG 283  
Db 1427 CAATAATTCAGATTCCTTAACAGAGTGGCAATACCAAGAGAGTATTCAGATCCA 1486  
QY 264 GAAGGCTATGATGCTTGAAGAGATGCTCCCTTACAGAGATTAAGAAATTTAGAT 323  
Db 1487 AGCTTCAAGAGTTTCTTCTGCTGACCTTATGACCAAGATATTCAGATCTACCTT 1546  
QY 324 ACCGAGCTATATATGCGAGAGGATTTATCTGAGAAAGGCTCTATACCTCTGAGAG 383  
Db 1547 TCCGATCAGATATGTCGCAACGAGACCTTGACCGTGAAGTGTCAAGTTTTCACAAAG 1606  
QY 384 GAGTTATCCCGGACTTCTGCGAGCTGTCATATCCCGGATATTTGAACCCGTTGA 443  
Db 1607 CTCTCTGTCAGACCTGTCAGAGCTTCTGTCATGATTCCTATTCACCTGTTAGA 1666  
QY 444 GTATTAAGATTAAGTCTGTTGAGAGGATTAAGTAAACAACCTTCCCTGAGCCCTT 503  
Db 1667 GATTAAGGAAACCACTTAAGTGAAGAGAGCTGTTCAACACCTTCCGCTTCTGAT 1726  
QY 504 TCAAGAAAGGCTATTCACACGTTTGCTGATGATGCTCCCTCAATAGAGCGGAAAG 563  
Db 1727 TCGTGAAGAGCTGCTGATCTATGAGAGTACCTCAATCCAAAGCTTCCGACTA 1786  
QY 564 TATTAAGACATTTCTCAACATCTTTTGAAGAGCTTCTTTGCGGCTGCTCAATC 623  
Db 1787 TAAAGAGCTTTTGAGATTTGACAGAGGCTTCTGCTCAATCTTCCAGAGAACTC 1846  
QY 624 CGAAAGAGAGAGATTTTGTGACCTGCTATAGTCTCT--GAGCTTGAAGAGTTCAC 680  
Db 1847 CTTGCCGATGCGAAGCTGCTGATTTGCTATGCAATCAAGAGAAATTAATCAAGTTCA 1906  
QY 681 ACCCTTGAATTTGAAGAAAGCGAACCAATTAATGAGAGGAGATTAAGAGCTTTAGA 740  
Db 1907 ACCATTCACAGTGAAGCGGAGAAATTCCTGCTTGGATTAAGAAATTCGCAAA 1966  
QY 741 GGTACTT 747  
Db 1967 AGTCTT 1973

RESULT 4

US-09-090-793-12  
Sequence 12, Application US/09090793  
Patent No. 6140486  
GENERAL INFORMATION:  
APPLICANT: Calgene, LLC  
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
of the reference gene, 131.01US  
CURRENT FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,650  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 12  
LENGTH: 40138  
TYPE: DNA  
ORGANISM: Vibrio marinus  
US-09-090-793-12

Query Match 6.1%; Score 46.2; DB 3; Length 40138;  
Best Local Similarity 52.9%; Pred. No. 0.00016;  
Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
QY 23 AGATAAAGCTGCTTTCTTCGGGAGAGCTGCAAGGGCATAGCCACATAGTGTGTTGA 82  
DB 14817 AAATGCGCTAGTCTCTTCGCGGTGGTGGAAAGTATGCTCATCTTGGTGTATTA 14876  
QY 83 AGCTATAAACAGAGCTCGGTATAAGGCTGAGGCTTTAAGCGGGTGAAGCGCGGCA 142  
DB 14877 ATACTGTAGAGCAAGTATAGCAAGTATGATGCGGGTACAGTGTGCTCTA 14936  
QY 143 TCGTTTGGTCTTTTATGCTCAGGCTACTCCCTGAAAGGATGTTGAGCCTTCTGA 202  
DB 14937 TGGTTGGTCACTTTATTGCTCAGGACTTGAGATTGATGACATTTTACATTTCTTC 14996  
QY 203 GGTAAA 209  
DB 14997 ATGTAAA 15003

RESULT 5  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKGT NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:

CLONE: pTZgpt-Fls  
US-08-232-463-14  
Query Match 5.3%; Score 40.4; DB 1; Length 7218;  
Best Local Similarity 7.4%; Pred. No. 0.0052;  
Matches 32; Conservative 206; Mismatches 192; Indels 0; Gaps 0;  
QY 3 GAGATTGAGGAAATTTGAGAGATATAACCTGCTTCTTCGGGAGGAGCTGCAAGGCGCAT 62  
DB 1454 GAGATAGAGAAATTTGTACERRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395  
QY 63 AGCCACATAGTGTGTTTGAAGCTATAACAGAGCTCGGTATAGGTTAGGGTGGCTTTAAG 122  
DB 1394 RRR 1335  
QY 123 CGGGGTAGCGCGGGGCAATCGTTTCGGTCTTTTATGCGCTCAGGCTACTCCCTGAAGG 182  
DB 1334 RRR 1275  
QY 183 GATGTTCAAGCCTTCTGAAGAGGGTAAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACTCT 242  
DB 1274 RRR 1215  
QY 243 GAAGGATTGATAGGTTGGGAGAGGCTATAAGATTCTTTCGAGGAAGTTCTCCCTACAG 302  
DB 1214 RRR 1155  
QY 303 GAGATAGAAAACTTGAGATACCGACTATATATGCGGACGAGATTATCTCGGGAAG 362  
DB 1154 RRR 1095  
QY 363 GGCTCTATACCTCTCGGAGGAGTTTAAATCCCGCACTTCTCGGAGCTGTGCAATTC 422  
DB 1094 RRR 1035  
QY 423 CGGCATATTT 432  
DB 1034 GGAATTAAT 1025

RESULT 6  
US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: TUBERCULOSIS  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 4.3%; Score 32.2; DB 4; Length 4403765;  
Best Local Similarity 52.6%; Pred. No. 52;  
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
QY 404 TCGGAGCTGTGCAATTCGCGCATATTTGAACCGCTTGAGTAAAGAAATTTACTGTGCG 463  
DB 370685 TCGCGGAGTTGAGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 370744

QY 464 TTGACGAGGTATAGTAAACACTTCCCGTTGAGCCCTTTCAGAAAGCGGTATTCCTCA 523  
DB 370745 GTGCGCGGTGTTCGCGATGCAACGTTTCGTTGCCGAGTTGAAAGCCGATTTTCCG 370804  
QY 524 CCGTTTGCGTTGA 536  
DB 370805 CTGCCCGAGTTGA 370817

RESULT 7  
US-09-103-840A-1  
Sequence 1, Application US/09103840A

Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 4.3%; Score 32.2; DB 4; Length 4411529;  
Best Local Similarity 52.6%; Pred. No. 52;  
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 404 TCGGAGAGTCGTGCAATTCCTCCGATTTGAAACCGTTGAGTATTAAGATTACTGCTCG 463  
DB 370628 TCGCCGAGTTAAAGAGCGGTGTGTTGCTGCGGTTTCCAGCCGAGATTGCG 370687  
QY 464 TTGACGAGGTATAGTAAACACTTCCCGTTGAGCCCTTTCAGAAAGCGGTATTCCTCA 523  
DB 370688 GTGCGCGGTGTTCGCGATGCAACGTTTCGTTGCCGAGTTGAAAGCCGATTTTCCG 370747  
QY 524 CCGTTTGCGTTGA 536  
DB 370748 CTGCCCGAGTTGA 370760

RESULT 8  
US-09-008-303-1/C  
Sequence 1, Application US/09008303  
Patent No. 6033889  
GENERAL INFORMATION:  
APPLICANT: HAN, YE SUN  
APPLICANT: YU, GYU YU  
APPLICANT: KIM, SUNG HOU  
APPLICANT: LIM, JAE HMAN  
APPLICANT: RYU, JAE RYON  
APPLICANT: CHOI, IN GEOL  
TITLE OF INVENTION: GENE SEQUENCE OF AQUIFEX PYROPHILUS  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE AND PROTEIN EXPRESSED IN ESCHERICHIA  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OHLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/008,303  
FILING DATE: 16-JAN-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 97-1140  
FILING DATE: 16-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: OHLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2901-0109-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-1000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2057 base pairs  
TYPE: nucleic acid  
STRANDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1014..1652  
US-09-008-303-1

Query Match 4.2%; Score 31.4; DB 3; Length 2057;  
Best Local Similarity 54.9%; Pred. No. 2.5;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 433 GAACCCGTTGAGTATTAAGATTACTGCTGTTGACGAGGTATTAACACTTCC 492  
DB 541 GAAGCCTTGAGAGGCAATATTAAGACCGTTGCGAGCTATTCCTGAAACTTGAG 482  
QY 493 GTGAGCCCTTTCAGAAAGCGGTATTCACCGTTGCGTTGATTCCTCC 545  
DB 481 GTAATGAGAGGAGGAGGATTAATACACCGTTGAGCTTGAAGTGGCTCC 429

RESULT 9  
US-09-484-970B-101/C  
Sequence 101, Application US/09484970B  
Patent No. 6426186  
GENERAL INFORMATION:  
APPLICANT: Jones, Karen A.  
APPLICANT: Volkmuth, Wayne  
APPLICANT: Walker, Michael G.  
TITLE OF INVENTION: BONE REMODELING GENES  
FILE REFERENCE: PB-0014 US  
CURRENT APPLICATION NUMBER: US/09/484,970B  
CURRENT FILING DATE: 2000-01-18  
NUMBER OF SEQ ID NOS: 172  
SOFTWARE: PERL Program  
SEQ ID NO 101  
LENGTH: 2409  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6426186 404040.2CBI  
NAME/KEY: unsure  
LOCATION: 11-13, 15-17, 25, 35, 1273, 1281, 1288, 2402  
OTHER INFORMATION: a, t, c, g, or other  
US-09-484-970B-101

Query Match 4.1%; Score 31; DB 4; Length 2409;  
Best Local Similarity 64.8%; Pred. No. 3.7;  
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 46 GGAGCTGCAAGGCGATAGCCCATAGTGTGTTGAAGCTATATAACGAGCTCGTATA 105  
 Db 876 GCACGACAAAGCGATTTCACATATGTTTGGAGTAATTAAGTAACCTCTGTATA 817  
 QY 106 AGGCTGAGGCG 116  
 Db 816 AAAATAAATGC 806

## RESULT 10

US-08-747-574-3/c  
 ; Sequence 3, Application US/08747574  
 ; Patent No. 6015939  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CALGENE, INC.  
 ; TITLE OF INVENTION: PLANT VDE GENES AND  
 ; TITLE OF INVENTION: METHODS RELATED THERETO  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Calgene, Inc.  
 ; STREET: 1920 Fifth Street  
 ; CITY: Davis  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 95616  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB  
 ; COMPUTER: Apple Macintosh  
 ; OPERATING SYSTEM: Macintosh 7.5  
 ; SOFTWARE: Microsoft Word 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/747,574  
 ; FILING DATE: No. 6015939ember 7, 1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/023,502  
 ; FILING DATE: August 6, 1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/006,315  
 ; FILING DATE: No. 6015939ember 7, 1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Donna E. Scherer  
 ; REGISTRATION NUMBER: 34,719  
 ; NAME: Carl J. Schwedler  
 ; REGISTRATION NUMBER: 36,924  
 ; REFERENCE/DOCKET NUMBER: 119-2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (916) 753-6313  
 ; TELEFAX: (916) 753-1510  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1555  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA to mRNA

Query Match 4.0%; Score 30.4; DB 3; Length 1555;  
 Best Local Similarity 53.3%; Pred. No. 4.6;  
 Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
 QY 582 CATCTTTGAGGAGCTTCTTTCTTGGCGTCAAACTCCGAAAGAAAGAGGTT 641  
 Db 311 CAGCTCTTTCAGCTCAITCTTGTATGTAATGGCAATGTCAATATCCCTTAGAGAA 252  
 QY 642 TTGTGACCTCGTTATAGTCTCTGAGCTTGGAGGTTACACCCCTTGATGTAGAAAGC 701  
 Db 251 TCCTGACCTGAATCCAGATAAAGACGTGAGGATCTCCACCAGTTGTTCTGAGATCAGC 192

## RESULT 11

## US-08-607-509-3

; Sequence 3, Application US/08607509  
 ; Patent No. 5876735  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED and BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/607,509  
 ; FILING DATE: 16-FEB-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.404C3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1867 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 117..1325

## US-08-607-509-3

Query Match 3.9%; Score 29.8; DB 2; Length 1867;  
 Best Local Similarity 48.5%; Pred. No. 8.1;  
 Matches 82; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
 QY 126 GGTGAGCGCGGGCAATCGTTTCGGTCTTTTATGCTCTCAGGCTACTCCCTCGAAGGAT 185  
 Db 641 GGTGCTCGACGAGGCTGATGATGCTCTCAGGCTTCGCGACCAAGATTTACGAGAT 700  
 QY 186 GTTCAGCCTTCTGAAAGGGTAAACTGGCTGAAGCTGTTTAAGTTCAAGCCACCTCTGAA 245  
 Db 701 CTTCGGCTTCTCGCGAAGGACATCCAGGTCGGCTCTTCTCCGCCACGATGCCGGAGGA 760  
 QY 246 GGGATTGATAGGTTGGGAGAGGCTATAAGATTCCTTGAGGAAGTTCTC 294  
 Db 761 GGTACTGGAGCTGACGAGAAGTTCTATCGCGACCCCGTGCGTATTCTC 809

## RESULT 12

US-08-634-642-3  
 ; Sequence 3, Application US/08634642  
 ; Patent No. 5879687  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF  
 ; TITLE OF INVENTION: PROTECTIVE IMMUNE RESPONSES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED and BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA

ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/634,642  
FILING DATE: 18-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.404C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 117..1325  
US-08-634-642-3

Query Match 3.9%; Score 29.8; DB 2; Length 1867;  
Best Local Similarity 48.5%; Pred. No. 8.1;  
Matches 82; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 126 GGTGAGCGCGCGGCAATCGTTGCTCTTTATGCTCAGGCTACTCCCTGAAGGAT 185  
|||  
DB 641 GGTGCTGACGAGGCTGATGATGCTGCTCAGGCTTGCGGACGAGATTACAGAT 700  
|||

QY 186 GTTCAGCCTTCTGAAGAGGTTAACTGCTGAAGCTGTTTAAGTTCAAGCCACTCTGAA 245  
|||  
DB 701 CTTCGCTTCTGCGGAGGACATCCAGTCCGCTCTTCTCCGCGACATGCCGAGGA 760  
|||

QY 246 GGGATTGATAGGTGGAGAGGCTATTAAGATTCTTGAGAAATTCTC 294  
|||  
DB 761 GGTACTGAGCTGACGAGAGAGATTCAATGCGCACCCCGTATTCTC 809  
|||

RESULT 13  
US-08-989-370-3  
Sequence 3, Application US/08989370  
Patent No. 6013268  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: SEED and BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/989,370  
FILING DATE: 12-DEC-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.404C5

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 117..1325  
US-08-989-370-3

Query Match 3.9%; Score 29.8; DB 3; Length 1867;  
Best Local Similarity 48.5%; Pred. No. 8.1;  
Matches 82; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 126 GGTGAGCGCGCGGCAATCGTTGCTCTTTATGCTCAGGCTACTCCCTGAAGGAT 185  
|||  
DB 641 GGTGCTGACGAGGCTGATGATGCTGCTCAGGCTTGCGGACGAGATTACAGAT 700  
|||

QY 186 GTTCAGCCTTCTGAAGAGGTTAACTGCTGAAGCTGTTTAAGTTCAAGCCACTCTGAA 245  
|||  
DB 701 CTTCGCTTCTGCGGAGGACATCCAGTCCGCTCTTCTCCGCGACATGCCGAGGA 760  
|||

QY 246 GGGATTGATAGGTGGAGAGGCTATTAAGATTCTTGAGAAATTCTC 294  
|||  
DB 761 GGTACTGAGCTGACGAGAGAGATTCAATGCGCACCCCGTATTCTC 809  
|||

RESULT 14  
US-08-602-359A-26/C  
Sequence 26, Application US/08602359A  
Patent No. 5942430  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFFIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,359A  
FILING DATE: February 16, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:  
LENGTH: 756 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
US-08-602-359A-26

Query Match 3.9%; Score 29.6; DB 2; Length 756;  
Best Local Similarity 53.4%; Pred. No. 5.7;  
Matches 62; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 516 TATTCCACCGTTTCGGTGTGATGCTTCCATAGACCGGAAAGGATATAAGAACAT 575  
DB 631 TCTTTTCGGAGTTTGAGCGGACCGCAAGAGAAGAGCTCTCAAAAGGATGTGAAGATGT 572  
QY 576 TCTTCACATCTTTTGGAGGAGCTTCTTTCTTGGGTCGCTCAAACTCCGAAAGA 631  
DB 571 TCTTTATATCTTTTCCGGCTCTATGGGAAGGACATCAAGCAACGGTGGGAATA 516

## RESULT 15

US-08-961-527-104  
Sequence 104, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6735 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-104

Query Match 3.9%; Score 29.6; DB 4; Length 6735;  
Best Local Similarity 59.5%; Pred. No. 19;  
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 193 CTTCTGAAGAGGGTAAACTGGCTGTAAGCTGTTTAAGTTCAGCCACCTCTGAAGGGATTG 252  
DB 5464 CTTTTCATATGCTAAACTTTGTGAGACGACGACGAGTTAAGGCCATCTATGAGACTATT 5523  
QY 253 ATAGGTGGGAGAGGCTATAGA 276

Db 5524 TAAATGAATGGCAAGTTACTAAGA 5547  
Search completed: June 19, 2003, 14:22:18  
Job time : 64 secs





SEQUENCE CHARACTERISTICS:  
LENGTH: 756 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 26;  
US-10-027-805-26

Query Match 100.0%; Score 756; DB 9; Length 756;  
Best Local Similarity 100.0%; Pred. No. 1.9e-247;  
Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TTGAGATTGAGGAAATTTGAAGAGATAAACTCGTTCTTCGCGAGAGCTGCAAGGGC 60
DB 1 TTGAGATTGAGGAAATTTGAAGAGATAAACTCGTTCTTCGCGAGAGCTGCAAGGGC 60
QY 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGCTGAGGGCTTTA 120
DB 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGCTGAGGGCTTTA 120
QY 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCGCTCAGGCTACTCCCTGAA 180
DB 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCGCTCAGGCTACTCCCTGAA 180
QY 181 GCGATGTTCAAGCTTCTGAAGAGGGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
DB 181 GCGATGTTCAAGCTTCTGAAGAGGGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
QY 241 CTGAAGGCAATGTAGGGTGGGAGAGGCTATAAGATTCTTGAAGAGTCTTCCCTTAC 300
DB 241 CTGAAGGCAATGTAGGGTGGGAGAGGCTATAAGATTCTTGAAGAGTCTTCCCTTAC 300
QY 301 AGGAGATAGAAAACCTTGAGATACCGAGTATATATGCGGACGGAATTTACTCGGGA 360
DB 301 AGGAGATAGAAAACCTTGAGATACCGAGTATATATGCGGACGGAATTTACTCGGGA 360
QY 361 AGGGCTCTATACCTCTCGAAGGGAGTTTAACTCCCGCACTTCTCGGAGCTGTGCAATT 420
DB 361 AGGGCTCTATACCTCTCGAAGGGAGTTTAACTCCCGCACTTCTCGGAGCTGTGCAATT 420
QY 421 CCGGCAATTTGAAACCGCTTGAGTATAAGAAATTAATGCTCGTTGACGGAGTATAGT 480
DB 421 CCGGCAATTTGAAACCGCTTGAGTATAAGAAATTAATGCTCGTTGACGGAGTATAGT 480
QY 481 AACAACTTCCCGTTCAGCCCTTTCAGGAAGGGGTATTCACACGTTTGGTGAATGC 540
DB 481 AACAACTTCCCGTTCAGCCCTTTCAGGAAGGGGTATTCACACGTTTGGTGAATGC 540
QY 541 CTTCCTATAGAGCCGAAAGGATATAAGAAACATTTCTACATCCTTTTGAAGGAGCTTC 600
DB 541 CTTCCTATAGAGCCGAAAGGATATAAGAAACATTTCTACATCCTTTTGAAGGAGCTTC 600
QY 601 TTCTTCCGTCGCTCAAACTCCGAAAGAGAAAGAGTGTGACCTCGTTATAGTT 660
DB 601 TTCTTCCGTCGCTCAAACTCCGAAAGAGAAAGAGTGTGACCTCGTTATAGTT 660
QY 661 CTTGAGCTTGAGGAGTTTCAACCCCTTGATGTTAGAAAGCGGCAATTAATGAGAGG 720
DB 661 CTTGAGCTTGAGGAGTTTCAACCCCTTGATGTTAGAAAGCGGCAATTAATGAGAGG 720
QY 721 GGATACATAAGGCTTAGAGGTACTTTCTGAATAG 756
DB 721 GGATACATAAGGCTTAGAGGTACTTTCTGAATAG 756
```

RESULT 2

US-10-027-804-26  
Sequence 26, Application US/10027804  
Publication No. US20030054530A1  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
MURPHY, Dennis

REID, John  
MAFFIA, Anthony  
LINK, Steven  
SWANSON, Ronald V.  
WARREN, Patrick V.  
KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/027,804  
FILING DATE: 21-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,359  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 756 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 26;  
US-10-027-804-26

Query Match 100.0%; Score 756; DB 9; Length 756;  
Best Local Similarity 100.0%; Pred. No. 1.9e-247;  
Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TTGAGATTGAGGAAATTTGAAGAGATAAACTCGTTCTTCGCGAGAGCTGCAAGGGC 60
DB 1 TTGAGATTGAGGAAATTTGAAGAGATAAACTCGTTCTTCGCGAGAGCTGCAAGGGC 60
QY 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGCTGAGGGCTTTA 120
DB 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGCTGAGGGCTTTA 120
QY 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCGCTCAGGCTACTCCCTGAA 180
DB 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCGCTCAGGCTACTCCCTGAA 180
QY 181 GCGATGTTCAAGCTTCTGAAGAGGGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
DB 181 GCGATGTTCAAGCTTCTGAAGAGGGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
QY 241 CTGAAGGCAATGTAGGGTGGGAGAGGCTATAAGATTCTTGAAGAGTCTTCCCTTAC 300
DB 241 CTGAAGGCAATGTAGGGTGGGAGAGGCTATAAGATTCTTGAAGAGTCTTCCCTTAC 300
QY 301 AGGAGATAGAAAACCTTGAGATACCGAGTATATATGCGGACGGAATTTACTCGGGA 360
DB 301 AGGAGATAGAAAACCTTGAGATACCGAGTATATATGCGGACGGAATTTACTCGGGA 360
QY 361 AGGGCTCTATACCTCTCGAAGGGAGTTTAACTCCCGCACTTCTCGGAGCTGTGCAATT 420
```

Thu Jun 19 17:30:37 2003

us-09-903-410-26.rmpb

Page 3

Db 361 AGGGCTCTATACCTCTCGGAGAGATTAAATCCCGCACTTCGCGAGCTGGCAATT 420  
Qy 421 CCGGCAATTTGAACCCGTTGAGTATAGAAATTAATCTGCTGTTGACGAGATAGT 480  
Db 421 CCGGCAATTTGAACCCGTTGAGTATAGAAATTAATCTGCTGTTGACGAGATAGT 480  
Qy 481 AACAACTTCCGTTGAGCCCTTTGAGAAAGCGGATTTCCACCGTTTGCGTTGATGTC 540  
Db 481 AACAACTTCCGTTGAGCCCTTTGAGAAAGCGGATTTCCACCGTTTGCGTTGATGTC 540  
Qy 541 CTTCCTATAGAGCCGAGAAAGATATTAAGAAACATTTCTTCAATCTTTGAGAGCTTC 600  
Db 541 CTTCCTATAGAGCCGAGAAAGATATTAAGAAACATTTCTTCAATCTTTGAGAGCTTC 600  
Qy 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGATTTTGTGACTGTTATAGT 660  
Db 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGATTTTGTGACTGTTATAGT 660  
Qy 661 CCTGAGCTTGAAGAGTTGACACCCCTTGATGTTGAAAAGGAGCAATTAATGAGAG 720  
Db 661 CCTGAGCTTGAAGAGTTGACACCCCTTGATGTTGAAAAGGAGCAATTAATGAGAG 720  
Qy 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756  
Db 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756

RESULT 3

US-09-903-410-26  
Sequence 26, Application US/09903410  
Patent No. US20020146799A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSEA CORPORATION  
APPLICANT: ROBERTSON, Dan  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFFIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald  
APPLICANT: WARREN, Patrick  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF  
FILE REFERENCE: DIVERSEA-180-2  
CURRENT APPLICATION NUMBER: US/09/903,410  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 09/382,242  
PRIOR FILING DATE: 1999-08-24  
PRIOR APPLICATION NUMBER: US 08/602,359  
PRIOR FILING DATE: 1996-02-16  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 26  
LENGTH: 756  
TYPE: DNA  
ORGANISM: Aquifex pyrophilus  
US-09-903-410-26

Query Match 100.0%; Score 756; DB 10; Length 756;  
Best Local Similarity 100.0%; Ered. No. 1.9e-247;  
Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGAGATGAGGAAATTTGAAGAGATAAACCCTGTTCTTGGGAGAGAGCTGCAAAAGGC 60  
Db 1 TTGAGATGAGGAAATTTGAAGAGATAAACCCTGTTCTTGGGAGAGAGCTGCAAAAGGC 60  
Qy 61 ATAGCCCATAGAGTGTGTTTGAAGACTATAAAGAGCTGATTAAGAGGAGGCTTTA 120  
Db 61 ATAGCCCATAGAGTGTGTTTGAAGACTATAAAGAGCTGATTAAGAGGAGGCTTTA 120  
Qy 121 AGCGGGGTAGAGCGCGGAGCATGCTTTCGCTTTTATGCTGCTAGCTATCCCGTGA 180  
Db 121 AGCGGGGTAGAGCGCGGAGCATGCTTTCGCTTTTATGCTGCTAGCTATCCCGTGA 180

Qy 181 GGGATGTTACGCTTTCTGAAGAGGTTAACTGCGTGAAGCTGTTTATGTTCAAGCACC 240  
Db 181 GGGATGTTACGCTTTCTGAAGAGGTTAACTGCGTGAAGCTGTTTATGTTCAAGCACC 240  
Qy 241 CTGAAGGATTAATGAGTGGTGGAGAGAGCTTATTAAGATTTCTTGAGAAAGTTCTCCCTTAC 300  
Db 241 CTGAAGGATTAATGAGTGGTGGAGAGAGCTTATTAAGATTTCTTGAGAAAGTTCTCCCTTAC 300  
Qy 301 AGGAGATTAAGAAATCTTGAGATACGACATATATATGCGGAGAGATTTATCTCGGGA 360  
Db 301 AGGAGATTAAGAAATCTTGAGATACGACATATATATGCGGAGAGATTTATCTCGGGA 360  
Qy 361 AGGCTCTATACCTCTGAGAAAGGATTTATCCCGCACTTCTCGGAGCTGTCGAATT 420  
Db 361 AGGCTCTATACCTCTGAGAAAGGATTTATCCCGCACTTCTCGGAGCTGTCGAATT 420  
Qy 421 CCGGCAATTTGAACCCGTTGAGTATAGAAATTAATCTGCTGTTGACGAGATATAGT 480  
Db 421 CCGGCAATTTGAACCCGTTGAGTATAGAAATTAATCTGCTGTTGACGAGATATAGT 480  
Qy 481 AACAACTTCCGTTGAGCCCTTTGAGAAAGCGGATTTCCACCGTTTGCGTTGATGTC 540  
Db 481 AACAACTTCCGTTGAGCCCTTTGAGAAAGCGGATTTCCACCGTTTGCGTTGATGTC 540  
Qy 541 CTTCCTATAGAGCCGAGAAAGATATTAAGAAACATTTCTTCAATCTTTGAGAGCTTC 600  
Db 541 CTTCCTATAGAGCCGAGAAAGATATTAAGAAACATTTCTTCAATCTTTGAGAGCTTC 600  
Qy 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGATTTTGTGACTGTTATAGT 660  
Db 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGATTTTGTGACTGTTATAGT 660  
Qy 661 CCTGAGCTTGAAGAGTTGACACCCCTTGATGTTGAAAAGGAGCAATTAATGAGAG 720  
Db 661 CCTGAGCTTGAAGAGTTGACACCCCTTGATGTTGAAAAGGAGCAATTAATGAGAG 720  
Qy 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756  
Db 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756

RESULT 4

US-10-027-805-29  
Sequence 29, Application US/10027805  
Patent No. US20020164725A1  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFFIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESSES:  
ADDRESS: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/027,805  
FILING DATE: 21-Dec-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,359  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 750 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRADEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 29:  
US-10-027-805-29

Query Match 52.6%; Score 397.8; DB 9; Length 750;  
Best Local Similarity 72.4%; Pred. No. 4e-125;  
Matches 516; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 TTGAGATTCAGGAATTTGAGAGATAAAGCTGCTTTCGCGAGGAGCTGCAAGGCG 60  
Db |||||  
QY 13 TTGAAATTCAGAGGTTTGAGAGGTTAACTTAGTTCTTTCGCGAGGAGCTGCAAGGTT 72  
Db |||||  
QY 61 ATAGCCCATATAGTGTGTTTGAAGCTATAAAGCTCGGTATAGGGGTGAGGCTTTTA 120  
Db |||||  
QY 73 ATGCCCATATAGTGTGTTTGAAGCTCTGGAAGAGCTCGGTATAAAGGTAAGAGGCTC 132  
Db |||||  
QY 121 AGCGGGTTCAGCGCGGCGCAATCGTTTCGCTCTTTTATGCTCAGGCTACTCCCTGAA 180  
Db |||||  
QY 133 AGCGGGTTCAGCGCTGAGGAGTATCGTTTCGCTCTTTTACGTCGCGGCTACCTCCGAC 192  
Db |||||  
QY 181 GGATGTTTCAGGCTTCGAGAGGAGTAACTGGCTGAGGCTGTTTAAAGTTCAGGCTACT 240  
Db |||||  
QY 193 GAGATGTTAAACTCTGTAAGAGAGTAACTGGCTCAAACTTTTAAAGTTCAAAACACGG 252  
Db |||||  
QY 241 CTGAAGGATTCATAGGTTGGAGAGGCTATAAGATTCCTTTCGAGGAGGTTCTCCCTTAC 300  
Db |||||  
QY 253 AAAATGGGCTTAATGGGTTGGAGAGGCTCGAGGTTTTTGGAAAAGAGCTCGGAGTT 312  
Db |||||  
QY 301 AGGAGATAGAAACTTTCGAGTACCGAGCTATATATGCGGAGCGGATTTTACTCGGA 360  
Db |||||  
QY 313 AAGGCTCGAAGACTGAACTACCAACCTATCTTTCGCGGAGTCTGTACCGGA 372  
Db |||||  
QY 361 AGGCTCTATACCTTCGGAAGGAGTTTAAATCCCGCACTTCGCGAGCTGTGCAATT 420  
Db |||||  
QY 373 AAGGCTCTTACTTCGCGAGAGTGACCTTAATTCGCGTGTCTTCGGAAGTTCTCCATA 432  
Db |||||  
QY 421 CCGGCAATTTTGAACCGCTTCAGTATAGAAATTAATTCGCTGTTGACGAGGTATAGTT 480  
Db |||||  
QY 433 CCGGCAATTTTGAACCGCTTCAGTATAGAAATTTTCTACTTTCGAGGAGTATAGTG 492  
Db |||||  
QY 481 AACACCTTCGCTTCAGGCTTCAGGAAAGCGGTATTCACCGGTTTCGCTGATGTC 540  
Db |||||  
QY 493 AACACCTTCGCTTCAGGCTTCAGGAAAGCTTCAGGAAAGCTTCAGGAAAGCTTCAGG 552  
Db |||||  
QY 541 CTTCCCATAGAGCCGGAAGGATATAAAGAACTTCCTTCACTCTTCCTTTCGAGGCTTC 600  
Db |||||  
QY 553 CTTCCCATAGAGCCGGAAGGATATAAAGAACTTCCTTCACTCTTCCTTTCAGGCTTC 612  
Db |||||  
QY 601 TTTCTTCGCTTCGCTTCAGGCTTCGGAAGGAGGTTTTCGCTTCGCTTCGCTTCGCTTC 660  
Db |||||  
QY 613 TTTCTTCGCTTCGCTTCAGGCTTCGGAAGGAGGTTTTCGCTTCGCTTCGCTTCGCTTC 672  
Db |||||  
QY 661 CTTGAGCTTCGAGGTTTCACCCCTTCGATGTTTGAAGAGCGGACCAATAAT 713  
Db |||||  
QY 673 CCTCCCTTCGAGGTTCTCTCTCTCTCGGAGCGTAAATAGGCGGAGGAGATTT 725  
Db |||||

RESULT 5  
US-10-027-804-29  
Sequence 29, Application US/10027804  
Publication No. US20030054530A1  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
MURPHY, Dennis  
REID, John  
MAFFIA, Anthony  
LINK, Steven  
SWANSON, Ronald V.  
WARREN, Patrick V.  
KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P. C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/027,804  
FILING DATE: 21-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,359  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5099  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 750 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRADEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 29:  
US-10-027-804-29

Query Match 52.6%; Score 397.8; DB 9; Length 750;  
Best Local Similarity 72.4%; Pred. No. 4e-125;  
Matches 516; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 TTGAGATTCAGGAATTTGAGAGATAAAGCTGCTTTCGCGAGGAGCTGCAAGGCG 60  
Db |||||  
QY 13 TTGAAATTCAGAGGTTTGAGAGGTTAACTTAGTTCTTTCGCGAGGAGCTGCAAGGTT 72  
Db |||||  
QY 61 ATAGCCCATATAGTGTGTTTGAAGCTATAAAGCTCGGTATAGGGGTGAGGCTTTTA 120  
Db |||||  
QY 73 ATGCCCATATAGTGTGTTTGAAGCTCTGGAAGAGCTCGGTATAAAGGTAAGAGGCTC 132  
Db |||||  
QY 121 AGCGGGTTCAGCGCGGCGCAATCGTTTCGCTCTTTTATGCTCAGGCTACTCCCTGAA 180  
Db |||||  
QY 133 AGCGGGTTCAGCGCTGAGGAGTATCGTTTCGCTCTTTTACGTCGCGGCTACCTCCGAC 192  
Db |||||  
QY 181 GGATGTTTCAGGCTTCGAGAGGAGTAACTGGCTGAGGCTGTTTAAAGTTCAGGCTACT 240  
Db |||||  
QY 193 GAGATGTTAAACTCTGTAAGAGAGTAACTGGCTCAAACTTTTAAAGTTCAAAACACGG 252  
Db |||||  
QY 241 CTGAAGGATTCATAGGTTGGAGAGGCTATAAGATTCCTTTCGAGGAGGTTCTCCCTTAC 300  
Db |||||



Db 14817 AAATTGCTTAGTCTCTTCTGGCGGTGGTGGAAAGTAATTGCTCATCTTGGTATTAA 14876  
Qy 83 AAGCTATATAACGAGCTGGTATTAAGGCTGAGGGCTTTAAGCGGGGTGAGCGCGGGCAA 142  
Db 14877 AATACCTGTTAGAGCAAGATATAGACGAATGTAATTCGGGTACAAAGTGTCTGCTTA 14936  
Qy 143 TCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAGGAGATGTTTCAGCCCTTCGAAGA 202  
Db 14937 TGGTTGCTGCACTTTTATGCTCAGGACTTGAGATTGATGACATTTTACAAATCTTCATCG 14996  
Qy 203 GGGTAAA 209  
Db 14997 ATGTAAA 15003

## RESULT 8

US-09-974-300-6748  
; Sequence 6748, Application US/09974300  
; Patent No. US20020146721A1

## GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Id Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: Expression  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6748  
; LENGTH: 467  
; TYPE: DNA  
; ORGANISM: Bacillus clausii

US-09-974-300-6748

Query Match 6.1%; Score 45.8; DB 10; Length 467;  
Best Local Similarity 44.4%; Pred. No. 4.7e-05;  
Matches 185; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

Qy 24 GATTAACCTCGTCTTTTCGGGAGGAGCTGCAAAAGGAGTATGTTAAGCGGGGTGAGCGCGGGCAAT 143  
Db 21 GATTGATGCGCCCTTGGGTCTGGCGGGCTAGAGGCTTTGCCCAATTTGGTGTAA 80  
Qy 84 AGCTATAACGAGCTCGGTATAAGGGTGAAGGCTTTAAGCGGGGTGAGCGCGGGCAAT 143  
Db 81 GACGCTTACAGAAAGCGGGGTTTCAATTTAGCTGTGTAGCAGCATGGGTGCAAT 140  
Qy 144 CGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAGGAGTGTTCAGCCTTCTGAAGAG 203  
Db 141 GGTGGGCTATGATGCGGTAGGCAATCGGTTGAACCATGGAAGTTTTCGCAACA 200  
Qy 204 GGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACTCTCAAGGGATTTGATAGGGTGA 263  
Db 201 TTTCAACGAAATATTACTTGGATTTTACAGTCTCAAAACAAGGGCTTATGCTGACA 260  
Qy 264 GAAGGCTATAAGATTCCTTGAGGAAGTCTCCCTTACAGGAGATAGAAAATCTGAGAT 323  
Db 261 AAGATTGAAGCTCATCTGTTTACTAGCTTAAAGACAAAGCTGAGTGCCTTTTCC 320  
Qy 324 ACCGAGTATATATCGGAGCGATTTATCTCCGGAAGGGCTCTATACCTCTCGGAAG 383  
Db 321 GCCAGTTCAAGTGGTAGCACTGATCTTTTAAAGTGGCCAAAAGTGTGATGTCAAAGG 380  
Qy 384 GAGTTTAATCCCGCACTTCTCGGAGCTGTGCAATTCGCGCATATTTCAACCCGT 440  
Db 381 AGACGTGCGCAAGCGGTTCGAGCAAGCTTGTGATTCAGGCATTTTGTCTCT 437

## RESULT 9

US-09-974-300-2350  
; Sequence 2350, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Id Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: Expression  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2350  
; LENGTH: 876  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis

US-09-974-300-2350

Query Match 4.7%; Score 35.8; DB 10; Length 876;  
Best Local Similarity 46.6%; Pred. No. 0.18;  
Matches 115; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 23 AGATAAACCTCGTCTTTTCGGGAGGAGCTGCAAAAGGAGTATGTTAAGCGGGGTGAGCGCGGGCAA 82  
Db 5 ACATCGACGGGTGTTTTTCAGCGCGGCATGAAGGGTTCGCTTGTGGAGCATACG 64  
Qy 83 AAGCTATAAACGAGCTCGGTATAAGGGTGAAGGCTTTAAGCGGGGTGAGCGCGGGCAA 142  
Db 65 AAGCGCTTGAAAAACGGGGCTTCCGATTTAAAGAGCTGGCGGAAACGAGCGCGGTTTCA 124  
Qy 143 TCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAGGAGTGTTCAGCCTTCTGAAGA 202  
Db 125 TCATGCTCTATTTATCGCGCGGCATATACAGCGGAGATCGGCAATGATGAGG 184  
Qy 203 GGTAAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACTCTCAAGGATTTGATAGGGTGG 262  
Db 185 AATTAATGAAGAGGAGCTCTCTTTCAGCGAGGTTTCACTCTTGCCTTTGAAATTTGCTGC 244  
Qy 263 AGAAGGC 269  
Db 245 AATGGGC 251

## RESULT 10

US-09-938-842A-1639  
; Sequence 1639, Application US/09938842A  
; Patent No. US20020160378A1

## GENERAL INFORMATION:

; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SAME, AND METHODS OF USE  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1639  
; LENGTH: 1062  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana

US-09-938-842A-1639

Query Match 4.6%; Score 35; DB 9; Length 1062;  
Best Local Similarity 49.2%; Pred. No. 0.37;  
Matches 92; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 196 CTGAAGAGGTAACCTGCTGAGCTTTTAAGTTACGCACTCTGAAGGATTTGATA 255  
DB 71 CCGCGCGCGCAAAATGGAAGATACCGCTTTGAGAGAGCTGCTTTCGAGGGATTCACG 130  
QY 256 GGGTGGGAGAGGCTATAGATTCTTGAAGAGTCTCCCTTACAGAGAAATGAAAA 315  
DB 131 GGTGAGAGAGTTCTTAACTGATCGGTCAAGTCAACACGAGAGAGATCAGA 190  
QY 316 CTGAGATACCGACGTATATATATGCGCGAGGATTATATCTCGGAGAGGCTCTATACCTC 375  
DB 191 CGAGATTAACCGCGGTGACTGACGTGCGCGTTAACAGCTTCAAGAGAGTCTATCTCTAC 250  
QY 376 TCGGAG 382  
DB 251 TCGGTAG 257

## RESULT 11

US-10-295-403-89  
Sequence 89, Application US/10295403  
Publication No. US20030101481A1

## GENERAL INFORMATION:

APPLICANT: Heard, Jacqueline  
APPLICANT: Riechmann, Jose Luis  
APPLICANT: Adam, Luc  
APPLICANT: Broun, Pierre  
APPLICANT: Pineda, Omaire  
APPLICANT: Reuber, Lynne  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Keddie, James  
APPLICANT: Zhang, James  
APPLICANT: Beutlo, Maria-Ines  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Fromm, Mike  
TITLE OF INVENTION: PLANT GENE SEQUENCES I  
FILE REFERENCE: M81-0003  
CURRENT FILING DATE: 2002-11-15  
PRIOR FILING DATE: US/09/394,519  
PRIOR FILING DATE: 1999-09-13  
PRIOR FILING DATE: 1998-09-22  
PRIOR FILING DATE: 1998-10-06  
PRIOR FILING DATE: 1998-11-17  
PRIOR FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 170  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 89  
LENGTH: 1390  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (35)..  
OTHER INFORMATION: G1022  
US-10-295-403-89

## Query Match

Best Local Similarity 4.6%; Score 35; DB 9; Length 1390;  
Matches 92; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 196 CTGAAGAGGTAACCTGCTGAGCTTTTAAGTTACGCACTCTGAAGGATTTGATA 255  
DB 105 CCGCGCGCGCAAAATGGAAGATACCGCTTTGAGAGAGCTGCTTTCGAGGGATTCACG 164

QY 256 GGGTGGGAGAGGCTATAGATTCTTGAAGAGTCTCCCTTACAGAGATAGAAAA 315  
DB 165 GTGTGAGAGAGTTCTTAACTGATCGGTCAAGTCAACACGAGAGAGTCTGAGA 224  
QY 316 CTGAGATACCGACGTATATATATGCGGAGTTTATCTCGGAGAGGCTCTATACCTC 375  
DB 225 CGAGATTAACCGCGGTACTGACGTGCGCGTTTAAAGCTTCAAGAGAGTCTATCTAC 284  
QY 376 TCGGAG 382  
DB 285 TCGGTAG 291

## RESULT 12

US-10-184-644-546  
Sequence 546, Application US/10184644  
Publication No. US20030044930A1

## GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zhenli  
TITLE OF INVENTION: ACIDS ENCODING THE SAME  
FILE REFERENCE: P3430R1C227  
CURRENT FILING DATE: 2002-06-28  
PRIOR APPLICATION REMOVED - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO: 546  
LENGTH: 458  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-184-644-546

## Query Match

Best Local Similarity 4.4%; Score 33.2; DB 9; Length 458;  
Matches 40; Conservative 113; Mismatches 252; Indels 0; Gaps 0;

QY 218 AGCTGTTAAGTTCAAGCCCTCTGAGGATGTATAGGTGGAGAAAGCTATAGAT 277  
DB 53 AATGATLTLRSKDPRAOHAAATKCTSPSTBLSTSRGATLLVATKASPAKLEBARALNQ 112  
QY 278 TCCCTTGAAGAGTCTCCCTTACAGAGATGAAAAAATGAGATACCGATATATAT 337  
DB 113 ALEMKRQKREKQKQKFMHAKMDPFDALTFEGIFSEBDKDIQADVLTYRALTLISPY 172  
QY 338 GCGGAGAGATTTATCTGCGGAGAGGCTCTATACCTCTCGAGAGGATTTATCCCG 397  
DB 173 HEKLVNRDRLPLVEIDRYSIISDKVKVMSIPKNSALRMEETVYHHYHTVA 232  
QY 398 CACTTCGCGAGCTGTCATTCCTCGGATATTTAACCCTTGAGTATTAAGTAACT 457  
DB 233 IEGNTLTLSEIRHILFTRAVVPKSLBEQNEVIGMIAAKKYLINTLVISIGVTSIDVLE 292  
QY 458 TGCCTGTTGACGAGATATATTAACACCTTCCCTTACGAGGCTTTACGAGAGCGGTA 517  
DB 293 IHRRLVGVDPVEAGRPRTTQVLVGHRIIPHPDVEKQOBFVQWLSSEANLHVEVEA 352  
QY 518 TTCCACCGTTTGCTGATGCTCTCCATAGAGCGGAGAAAGATATTAAGAACATTC 577  
DB 353 ALAHYKLVYIHPFDIGNRTSLMLMLILMQAGYPTITTRKGRSDYHYHLEAABGDVR 412  
QY 578 TTCAATCTTTTGAAGAGCTTCTTCTGCGGTCCGCTCAACT 622



CURRENT FILING DATE: 2001-06-25  
PRIOR APPLICATION NUMBER: US 60/213,848  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: US 60/214,087  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: US 60/258,692  
PRIOR FILING DATE: 2000-12-29  
NUMBER OF SEQ ID NOS: 875  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 66  
LENGTH: 2000  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-887-576-66

Query Match 4.3%; Score 32.8; DB 10; Length 2000;  
Best Local Similarity 52.1%; Pred. No. 3;  
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY	35	TTCTTTCGGAGAGAGCTGCAAGGCGCATAGGTCCTTTTGAAGCTATAACG	94
DB	188	TTTTTTTGTATATGACGATGCAAAAGCCCAAGGTAGCTGAAACATATATGAT	247
QY	95	AGCTGGTATAGGTTGAGGCTTTAAGCGGGTGAAGCCCGGGCAATGTTGGCT	154
DB	248	ATCAAAAATTAATGTTGATGATTTGATCATAGTTGACAAAGTGTTTACGTATATATTT	307
QY	155	TTTATGCTCAGGCTACTCC	174
DB	308	TTTACATTCAGGCTCTTAC	327

Search completed: June 19, 2003, 15:01:24  
Job time: 123 secs



Thu Jun 19 17:30:40 2003

us-09-903-410-26.rst

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 13:31:50 ; Search time 1077 Seconds

(without alignments)  
11368.420 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756

Sequence: 1 ttagagattgaggaatttga.....ttagagttacttctgaatag 756

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_ges:\*  
18: em\_ges\_hum:\*  
19: em\_ges\_inv:\*  
20: em\_ges\_pln:\*  
21: em\_ges\_vit:\*  
22: em\_ges\_fun:\*  
23: em\_ges\_mam:\*  
24: em\_ges\_mus:\*  
25: em\_ges\_other:\*  
26: em\_ges\_pro:\*  
27: em\_ges\_fod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	47.2	6.2	392	12	BF657575
2	41.8	5.5	470	13	BM107729
3	40.8	5.4	456	10	AV397710
4	40.8	5.4	657	13	BI875128
5	40.8	5.4	661	14	BO816618
6	40.8	5.4	713	14	BO817181

C	7	40.2	5.3	460	17	AZ517865
C	8	40	5.3	807	12	BG784079
C	9	39.8	5.3	459	17	AQ332328
C	10	39.8	5.3	465	17	AQ335962
C	11	38.6	5.1	510	17	AQ600815
C	12	38.6	5.1	1010	17	CNS01678
C	13	38.4	5.1	361	17	AZ049266
C	14	37.6	5.0	1244	12	AZ924693
C	15	37.6	5.0	1244	12	BF678161
C	16	35.8	4.7	698	17	BO808806
C	17	35.6	4.7	1101	17	CNS0000D1
C	18	35.2	4.7	767	12	BG845295
C	19	35	4.6	508	14	K30038
C	20	35	4.6	528	10	AV831633
C	21	35	4.6	569	9	A1992658
C	22	35	4.6	1101	13	BM479230
C	23	34.8	4.6	449	14	BO797080
C	24	34.8	4.6	463	17	AZ158907
C	25	34.8	4.6	589	14	BO785148
C	26	34.8	4.6	789	14	BO786737
C	27	34.8	4.6	816	9	AA789057
C	28	34.6	4.6	668	17	AO531614
C	29	34.4	4.6	461	10	AV723426
C	30	34.2	4.5	542	17	AO672914
C	31	34.2	4.5	633	17	AG154803
C	32	34	4.5	361	17	BH760289
C	33	34	4.5	408	9	A1164222
C	34	33.8	4.5	385	12	BE935559
C	35	33.8	4.5	426	14	BO143125
C	36	33.8	4.5	419	14	BO143278
C	37	33.8	4.5	596	17	AZ010196
C	38	33.8	4.5	619	17	FR0044642
C	39	33.8	4.5	668	17	AG077885
C	40	33.8	4.5	812	17	AG103741
C	41	33.8	4.5	898	17	CNS000QC
C	42	33.6	4.4	356	17	AQ279281
C	43	33.6	4.4	979	17	CNS06Y86
C	44	33.4	4.4	298	10	AM298975
C	45	33.4	4.4	451	10	AM147194

# ALIGNMENTS

AZ517865 RPT-11-8  
BG784079 SEAMC004  
AQ332328 HS\_5007\_A  
AQ335962 HS\_5015\_B  
AQ600815 HS\_5318\_A  
AL107180 Drosophila  
AZ049266 GSSBR1081  
AZ924693 4906\_1455  
BF678161 602084827  
BO808806 RPT-24-3  
AL065414 Drosophila  
BG845295 1024009P1  
K30038 12643\_Lambda  
AV831633 AV831633  
A1992658 701558595  
BM479230 AGENCOURT  
BO797080 EST\_6018  
AZ158907 SP\_0061\_B  
BO785148 EST\_4086  
BO786737 EST\_5675  
AA789057 ag41B12.s  
AO531614 RPT-11-3  
AV723426 AV723426  
AO672914 HS\_5460\_A  
AG154803 pan trogl  
BH760289 BMBAC112A  
A1164222 A057P53U  
BE935559 WR2-MT012  
BO143125 fmb1c.DKO  
BO143278 fmb1c.DKO  
AZ010196 RPT-23-3  
AL132134 Fugu rubr  
AG077885 pan trogl  
AG103741 pan trogl  
AL076805 Drosophila  
AQ279281 CITBI-E1  
AL420700 T3 end of  
AM298975 180 MARC  
AM147194 dal5c07.x

RESULT 1  
BF657575  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF657575 392 bp mRNA linear EST 20-DEC-2000  
OV2.23.H05.bl\_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA  
BF657575  
BF657575.1 GI:11922709  
EST.  
Sorghum.  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Sorghum.  
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
,L.H.  
An EST database from sorghum: ovaries of varying immature stages  
unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Seq primer: JBN REV  
High quality sequence stop: 392  
POLYA-No.  
Location/Qualifiers

# FEATURES

```

source
1..392
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 2 (OV2)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      79 a   94 c   104 g   115 t
ORIGIN
Query Match      6.2%; Score 47.2; DB 12; Length 392;
Best Local Similarity 56.4%; Pred. No. 0.0028;
Matches 88; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 340 GCGACGATTTATCTCGGAGGCTCTATACCTTCGGAAGGAGTTTATCCCGCA 399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 GCCACCAATTTAAGTACGGGAGCGTGAATTATGTTTACTGAAGCGATCTCCATCTT 279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 CTTCCTCGGCGAGCTGTGCAATTCGCGCATATTTCAACCGCTTGAGTATAGTAATTCTG 459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 ATTCCGCGCATATGAGTATTCAGGACTCATGCACTGTTCACATACGGCTACTGG 339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 CTCTGTTGAGGAGGTATAGTTAAACAACCTTCCCGTT 495
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 CTGGTTGATGGAGCAGTCGTTAAACCAATTCCTATT 375
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
BM107729/c
LOCUS BM107729 470 bp mRNA linear EST 01-JUL-2002
DEFINITION c01603 cDNA Century library Glycine max cDNA clone c01e03 5', mRNA
sequence.
ACCESSION BM107729
VERSION BM107729.1 GI:21637574
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 470)
AUTHORS Matthews,B.F., Khan,R., MacDonald,M. and Beard,H.
TITLE Resistance mechanisms in soybean: Gene expression profile at an
early stage of soybean cyst nematode invasion
JOURNAL Unpublished (2001)
COMMENT Contact: Alkharouf, N.W.
Soybean Genomics and Improvement Laboratory (SGIL)
US Department of Agriculture (USDA), ARS, PSI
Bldg.006, Rm 118, 10300 Baltimore Ave., Beltsville, MD 20705-2350,
USA
Tel: 301 504 5750
Fax: 301 504 5728
Email: alkharouf@ars.usda.gov.

FEATURES
source
1..470
Location/Qualifiers
/organism="Glycine max"
/cultivar="Century"
/db_xref="taxon:3847"
/clone="c01e03"
/clone_lib="cDNA Century library"
/tissue_type="cotyledons"
/dev_stages="Eight days post germination"
/note="Vector: pBluescript SK-; cDNA clones from mRNA
extracted from soybean cv. Century cotyledons 8 days post
germination."
BASE COUNT      140 a   97 c   86 g   144 t   3 others
ORIGIN
Query Match      5.5%; Score 41.8; DB 13; Length 470;
Best Local Similarity 50.3%; Pred. No. 0.14;

```

```

Matches 100; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 477 AGTTAACAACTTCCCGTTGAGCCCTTTCAGGAAGCGGTATTCACCGCTTTGGGTTGA 536
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 ATTTAACAACTCTCATTTGGGTGGCAATTCAGGAGGATTTTGTCTCAAGTTCAGCTTGA 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 537 TGTCTTCCCATAGAGCCGGAAGATATAAGAACATCTTCACATCCTTTTGGAGGAG 596
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 TTAATATCAAGCGGACCAATGTGTGGAAGACATGAATCTGATTAATTGCTGAGAT 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 597 CTTCCTTTTTCGCGTCCGCTCAAACTCCGAAAAGAGAAAGGAGTTTGTGACCTCGTTAT 656
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 CCAATTCCTGAGGTAGCCAAATCACCAAATGTNTCAGGAAGTCTTTTGGAGTCACTGA 216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 657 AGTTCTTGAGCTTGAGGAG 675
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 AGTTGCTGCTCAGGTTGAG 197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AV397710/c
LOCUS AV397710 456 bp mRNA linear EST 23-APR-2002
DEFINITION AV397710 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
cDNA clone CL49b11_r 5', mRNA sequence.
ACCESSION AV397710
VERSION AV397710.1 GI:6551926
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 456)
AUTHORS Asamizu,B., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE A large scale structural analysis of cDNAs in a unicellular green
alga. Chlamydomonas reinhardtii. I. Generation of 3433
non-redundant expressed sequence tags
JOURNAL DNA Res. 6 (6), 369-373 (1999)
MEDLINE 20152988
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakam@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..456
/organism="Chlamydomonas reinhardtii"
/db_xref="taxon:3055"
/clone="CL49b11_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stages="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      101 a   142 c   136 g   77 t
ORIGIN
Query Match      5.4%; Score 40.8; DB 10; Length 456;
Best Local Similarity 59.5%; Pred. No. 0.29;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 460 CTCGTTGACGGAGTATAGTTACACCTTCCCGTTGAGCCCTTTCAGGAAGCGGTATT 519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 CTTCCTTGGCGAGATGGCGGAGACGCTCCTCTGTTGAGGCCCTTGGGGAAGTGTCCGA 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 CCCACCGCTTTCGCTGATGTCTTCCCATAGAGCCGGAAGATATAAGAACAT 575
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 CTCAATGATGTCCTTGAAGCCATACCTTGTAGGCGCGCTCCAGGATCTTCTGAATAT 165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
BI875128/c
LOCUS BI875128 657 bp mRNA linear EST 11-OCT-2001
DEFINITION 963122A09.y1 C. reinhardtii CC-1690, Stress condition I, normalized

```

ACCESSION	Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
VERSION	B1875128
KEYWORDS	B1875128.1 GI:16073132
SOURCE	EST.
ORGANISM	Chlamydomonas reinhardtii.
REFERENCE	Chlamydomonas reinhardtii.
AUTHORS	Chlamydomonas reinhardtii.
TITLE	Bukatyra; Vitidiplantae; Chlorophyta; Chlorophyceae; Volvocales;
JOURNAL	Chlamydomonadaceae; Chlamydomonas.
COMMENT	1 (bases 1 to 657)
	Grossman,A., Chang,C.-W., Davies,J.J., Harris,E., Hauser,C., Lefebvre,
	P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
	Analyses of the Chlamydomonas reinhardtii Genome: A Model,
	Unicellular System for Analyzing Gene Function and Regulation in
	Vascular Plants. Project: 963
	Unpublished (2001)
	Contact: Charles Hauser
	DCMB Box 91000
	Duke University
	Durham, NC 27708-1000
	Tel: 919 613 8159
	Fax: 919 613 8177
	Email: chauser@duke.edu.
FEATURES	Location/Qualifiers
source	1..657
	/organism="Chlamydomonas reinhardtii"
	/strain="CC-1690 wild type mt+ 21gr"
	/db xref="taxon:3055"
	/clone_1lb="C. reinhardtii CC-1690, stress condition I,
	normalized, Lambda zap II"
	/note=Vector: pBluescript II SK-, Site 1: EcoRI; Site 2:
	XhoI; This library, constructed by John Davies and Jeffrey
	Mcdormott, combines cDNAs from CC-1690 cells grown to
	mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
	1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
	4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
	purified from each sample, pooled and cDNA synthesized.
	The cDNA was directionally cloned into lambda zap II
	(Stratagene) in the EcoRI (5') and XhoI (3') sites.
	pBluescript II SK- plasmids were excised from the lambda
	zap clones by superinfection with EXASist (Stratagene)
	phage. The library was normalized using method 4 described
	in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT	143 a 198 c 191 g 125 t
ORIGIN	
Query Match	5.4%; Score 40.8; DB 13; Length 657;
Best Local Similarity	59.5%; Pred. No. 0.36; Indels 0; Gaps 0;
Matches	69; Conservative 0; Mismatches 47;
OY	460 CTCGTTGACGAGATTAAGTAAACAACCTCCCGTTGAGCCCTTCGAGAAGCGGTATT 519
DB	383 CTTCTTGCGGAGATAGCGCGGAGACACGCCTCGTTGAGGCCCTTGCGGAAGTGGCA 324
OY	520 CCNACGTTTGCGTTGATGTCCTTCCCATAGAGCGGAAAAGATATAAGAACAT 575
DB	323 CTCAATGATGCTTGTAAGCCACTACTGTTGAGGCGCGGTCAAGATCTTCGAAATAT 268
RESULT 5	
LOCUS	B0816618 661 bp mRNA linear EST 01-AUG-2000
DEFINITION	1030058H1.Y1 C. reinhardtii CC-1690, Deflagellation (normalized),
ACCESION	B0816618
VERSION	B0816618
KEYWORDS	Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
SOURCE	B0816618.1 GI:22065882
ORGANISM	EST.
	Chlamydomonas reinhardtii.
	Chlamydomonas reinhardtii.
	Eukaryotes; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
	Chlamydomonadaceae; Chlamydomonas.
	1 (bases 1 to 661)
	Grossman,A., Chang,C.-W., Davies,J.J., Harris,E., Hauser,C., Lefebvre,
	P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.

**TITLE**  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1030

**JOURNAL COMMENT**  
Unpublished (2002)  
Contact: Charles Hauser  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

**FEATURES**

**source**  
location/Qualifiers  
1..661  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-1690 wild type mc+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_id="C. reinhardtii CC-1690, Deflagellation  
(normalized)", lambda zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; Deflagellation library, constructed by John Davies  
and Jeffrey McDermott, combines cDNAs from CC-1690 cells  
which had been re-synthesizing flagella for 15, 30 and 60  
min after being deflagellated by pH shock. PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda zap II  
(Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
zap clones by superinfection with Exsist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al., (1996) Genome Research 6: 791-806."

**BASE COUNT**  
147 a 197 c 188 g 129 t

**ORIGIN**

Query Match 5.4%; Score 40.8; DB 14; Length 661;  
Best Local Similarity 59.5%; Pred. No. 0.36;  
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 460 CTCCTTGAACGAGGTATGTTAAACAACCTTCGGTTGACCCTTTCAGAGAAAGCGATT 519  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 411 CTTCTTGACGGAGAATGAGGACGAGACGTCCTGTTGAGACCCCTTGAGAAAGTCCGA 352  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 520 CCCACCGTTTCGGTGAATGTCCTCCCATAGACCAGAAAGGATTAAGACAT 575  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 351 CTCATGATGCTTGAAGCAACTCTTGTAAGGCCCGCTCACAGATCTCTGAATAT 296  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

**RESULT 6**  
B0817181/c 713 bp mRNA linear EST 01-AUG-2002  
LOCUS 1030062012.x1 C. reinhardtii CC-1690, Deflagellation (normalized),  
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION B0817181  
VERSION B0817181.1 GI:22066838  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.  
1 (bases 1 to 713)  
Grossman A., Chang C.-H., Davies J., Harris B., Hauser C., Lefebvre  
P., McDermott J.P., Shigeri J., Slightow C. and Stern D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1030  
Unpublished (2002)  
Contact: Charles Hauser  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
location/Qualifiers

**JOURNAL COMMENT**

**FEATURES**

ps 0;

QY 51 TGCMAAGGCGATACCCGACATAGTGTGTTGAAAGCTATATAACGAGCTCGTATTAAGGGT 110  
|||||  
Db 44 TGCMAAGGCGATATATCCAGCTCGGACATCAAGAGAGTACGATCCCAT 103  
|||||  
QY 111 GAGGCGTTTAAAGCGGGGTGAGCGCGCGGAGCATCGTTTGGTCTTTATAGCC 162  
|||||  
Db 104 CGACATGTCGGCGGGGTGAGCATCGGGGCCCTCATCTCGGCCCTTACGCC 155  
|||||  
RESULT 9  
AO332328 459 bp DNA linear GSS 06-MAR-1999  
LOCUS HS\_5007\_A2\_H07\_T7 RPC111 Human Male BAC library Homo sapiens  
DEFINITION genomic clone Plate=583 Col=14 Row=O, DNA sequence.  
ACCESSION AO332328  
VERSION AO332328.1 GI:4129955  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 459)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 583 row: O column: 14  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 459.  
FEATURES  
source  
1..459  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=583 Col=14 Row=O"  
/clone\_1ib="RPC111 Human Male BAC library"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; RPC111 Human Male BAC library"  
BASE COUNT 147 a 78 c 74 g 159 t 1 others  
ORIGIN  
Query Match 5.3%; Score 39.8; DB 17; Length 459;  
Best Local Similarity 49.8%; Pred. No. 0.58;  
Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 199 AAGAGGTTAACTGCTGAAGCTGTTAAGTTCAGCCACTCTGAAGGAGTGTATAGG 258  
|||||  
Db 272 AAAAAGGTCAGAGTCAATTCATTTTCAGTTTAACTATATTTTAAAAAGAAATGA 213  
|||||  
QY 259 TGGAGAGGCTTAAAGTCTTGAAGAGTTCCTCCCTTACGAGATAGAAAACCT 318  
|||||  
Db 212 TTTGGGTTAAAAATTAAGATGAGCTTATCTCTCCCTTACAGTAAGACGAAACT 153  
|||||  
QY 319 GAGATACCGAGCTATATATGCGGAGGATTTATCTCGGAGAGGCTCTATACCTCG 378  
|||||  
Db 152 GAGCTGAAAACACATAGAGCTGTACTCATTTCTATTTCACAAAGAACTGTCTACTGCT 93  
|||||  
QY 379 GAAGGAGTTTAATCCCGCACT 401  
|||||  
Db 92 GTATGCAAAATATATCTCTCAGT 70  
|||||

RESULT 10  
AO335962 465 bp DNA linear GSS 06-MAR-1999  
LOCUS HS\_5015\_B1\_E07\_S66E RPC111 Human Male BAC library Homo sapiens  
DEFINITION genomic clone Plate=591 Col=13 Row=J, DNA sequence.  
ACCESSION AO335962  
VERSION AO335962.1 GI:4142941  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 465)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 591 row: J column: 13  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 465.  
FEATURES  
source  
1..465  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=591 Col=13 Row=J"  
/clone\_1ib="RPC111 Human Male BAC library"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; RPC111 Human Male BAC library"  
BASE COUNT 149 a 76 c 76 g 162 t 2 others  
ORIGIN  
Query Match 5.3%; Score 39.8; DB 17; Length 465;  
Best Local Similarity 49.8%; Pred. No. 0.59;  
Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 199 AAGAGGTTAACTGCTGAAGCTGTTAAGTTCAGCCACTCTGAAGGAGTGTATAGG 258  
|||||  
Db 246 AAAAGGCTCAGAGTCAATTCATTTTCAGTTTAACTATATTTTAAAAAGAAATGA 187  
|||||  
QY 259 TGGAGAGGCTTAAAGTCTTGAAGAGTTCCTCCCTTACGAGATAGAAAACCT 318  
|||||  
Db 186 TTTGGGTTAAAAATTAAGATGAGCTTATCTCTCCCTTACAGTAAGTACGAAAACCT 127  
|||||  
QY 319 GAGATACCGAGCTATATATGCGGAGGATTTATCTCGGAGAGGCTCTATACCTCG 378  
|||||  
Db 126 GAGCTGAAAACACATAGAGCTGTACTCATTTCTATTTCACAAAGAACTGTCTACTGCT 67  
|||||  
QY 379 GAAGGAGTTTAATCCCGCACT 401  
|||||  
Db 66 GTCTGCAAAATATATCTCTCAGT 44  
|||||  
RESULT 11  
AO600815 510 bp DNA linear GSS 10-JUN-1999  
LOCUS HS\_5318\_A2\_C09\_T7A RPC1-11 Human Male BAC library Homo sapiens  
DEFINITION genomic clone Plate=894 Col=18 Row=E, DNA sequence.



VERSION	AZ049266.1 GI:7273181
KEYWORDS	GSS.
SOURCE	Brucella melitensis biovar Abortus.
ORGANISM	Brucella melitensis biovar Abortus Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; Brucella. 1 (bases 1 to 361)
REFERENCE	Sanchez,D.O., Zandamoni,R.O., Cravero,S., Verdun,R.E., Pierrou,E., Paccio,P., Diaz,G., Lanzavecchia,S., Agüero,F., Frasch,A.C.C., Andersson,S.G.E., Rossetti,O.L., Grau,O. and Ugalde,R.A. Gene discovery through genomic sequencing of Brucella abortus Infect. Immun. 69 (2), 865-868 (2001) 21101034
TITLE	Contact: Zandamoni, R.
JOURNAL	Centro de Investigacion en Ciencias Agropecuarias (CITA)
MEDLINE	Instituto Nacional de Tecnologia Agropecuaria (INTA) C.C. 25 (1712)
COMMENT	Caelejar. Buenos Aires, Argentina Tel.: 5411-4621-3316/1683 Fax: 5411-4481-1316 Email: zandamoni@inta.gov.ar Class: shotgun. Location/Qualifiers 1..361 /organism="Brucella melitensis biovar Abortus" /strain="S-2308" /db_xref="taxon:235" /clone="B23" /clone_1fb="Sheared genomic library" /note="Vector: pBluescript SK(-) (STRAGENE); Genomic DNA was mechanically sheared, blunt ended, and size-fractionated by agarose gel electrophoresis. Fragments between 1.5-3 kb were recovered and ligated to the EcoRV site of the pBluescript SK (-) vector." BASE COUNT 88 a 112 c 97 g 63 t 1 others ORIGIN
FEATURES	
SOURCE	
Query Match	5.1%; Score 38.4; DB 17; Length 361;
Best Local Similarity	47.8%; Pred. No. 1.4;
Matches 111; Conservative	0; Mismatches 121; Indels 0; Gaps 0;
Oy	308 TAGAAAACTGATGCATACCGACGTATATATATGGCGACGATTATTAATCTCGGAAGGGCTC 367
Dd	245 TGGAAAGCTGAAAATTCCCATGACATTACGGCGTGTGATTTTATATGGCGCGGAGA 186
Oy	368 TATACTCTCGAAGAGGAGTTTAATCCCCGCACTTCGGGACGCTGCCAATTCGCCGCA 427
Dd	185 TTCAATATTAAGCATGTGATTTACGCTCGGCATAGTGCGTCCGCGCATTCACCAG 126
Oy	428 TATTGAACCCGTTGAGTATTAAGATTACTGCTCTGTTGAGGAGGATATAGTTAACACC 487
Dd	125 TTTTTCGCCCGGTGGCGGCAATGGCCGCATTTGATGATGGCGGCGTGYTCATCCGG 66
Oy	488 TTCGGTGGAGCCCTTTCAGGAAGACGGATATCCACCGTTTGCGTGTATG 539
Dd	65 TGCGCTGACCTCTTTTCGACAGGCGGATATGTAATGGCATGATG 14
RESULT 14	
AZ924693/c	462 bp DNA linear GSS 01-APR-2001
LOCUS	4906.id55b14.s1 Saccharomyces mikatae IFO 1815 Saccharomyces
DEFINITION	mikatae genomic clone 4906.id55b14.s1, DNA sequence.
ACCESSION	AZ924693
VERSION	AZ924693.1 GI:13495592
KEYWORDS	GSS.
SOURCE	Saccharomyces mikatae.
ORGANISM	Saccharomyces mikatae. Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. 1 (bases 1 to 462) Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish W.R., Waterston,R.H. and Johnston,M. Surveying Saccharomyces genomes to identify functional elements by
REFERENCE	
AUTHORS	
TITLE	

JOURNAL COMMENT	comparative DNA sequence analysis unpublished (2001)
REFERENCE SOURCE	Contact: Johnston M Department of Genetics Washington University Medical School Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA Tel: 314 362 2735 Fax: 314 362 7855 Email: mj@genetics.wustl.edu Class: random plasmid subclone. Location/Qualifiers 1..462 /organism="Saccharomyces mikatae" /strain="IFO 1815" /db_xref="taxon:114525" /clone="4906.id5b14.81" /clone_1lb="Saccharomyces mikatae IFO 1815" /note="Random genomic sequence"
BASE COUNT ORIGIN	142 a 107 c 72 g 141 t
Query Match	Best Local Similarity 5.0%; Score 37.8; DB 17; Length 462; Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Dy	10 AGAATTTTGAAGAGATAAACCTGCTTCCTTCGGAGAGACTGCAGGACATACCAC 69       
Dd	266 AGGAATATCCGTCCGACCGCTTAGTCTTAGTGAGAGGTGGTATTGCTTTGCCAC 207       
Oy	70 ATAGCTTTTGAAGCTATPAAGACTCCTGCTTAAGGTGAGCGCTTAAAGCGGGTG 129       
Dd	206 ATTGATGCTTCGTGTAGATTATTTGAACCTAGATTATTAACCTAGAAGATCAAGCGTAGC 147       
Oy	130 AGCGCCGGGCGCATCGT 146       
DB	146 AGTCCCGGTGCATTTGT 130       
RESULT 15	Bf678161 1244 bp mRNA linear EST 21-DEC-2000
Bf678161 LOCUS	60208482/P1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4248967 5', mRNA sequence.
DEFINITION	Bf678161 Bf678161.1 GI:11952056
ACCESSION VERSION KEYWORDS	EST. human.
SOURCE ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE AUTHORS TITLE JOURNAL COMMENT	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E.B. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LNLN at: http://image.llnl.gov Plate: LNCM1068 row: f column: 08 High quality sequence start: 6 High quality sequence stop: 88. Location/Qualifiers 1..1244 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4248967" /clone_1lb="NIH_MGC_83" /lab_host="DH10B (TI phage-resistant)"
FEATURES source	

/notes: Organ: prostate; Vector: pDMR-LIB (Clontech);  
Site 1: SfiI (ggcgcctcgcc); Site 2: SfiI (ggccattatggcc  
); 5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dT(30)BN-3'  
(where B = A, C, G, or T). Average  
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT	496 a	220 c	254 g	274 t
ORIGIN				
Query Match	5.0%; Score 37.6; DB 12; Length 1244;			
Best Local Similarity	52.6%; Pred. No. 5;			
Matches	82;	Conservative	0;	Mismatches 74; Indels 0; Gaps 0;
Qy	412	TGTGCAATCCCGGCAATTTTGAACCGTGTAGTATAAGAAATTAATCTTGCTCGTTGACGGA	471	
Db	133			
Qy	472	GGTATAGTTAACCAACCTTCCCGTTGAGCCCTTTCAGGAAAGCGGTATCCCAACCGTTTGC	531	
Db	193			
Qy	532	GTTGATGTCCTTCCCATAGACCGGAAAGGATATA	567	
Db	253	GCAGATGGTGAACCATAGGGTATTGGAATATAAA	288	

Search completed: June 19, 2003, 14:21:01  
Job time : 1081 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 13:05:45 ; Search time 1449 Seconds

(without alignments)  
15184.073 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756 1 ttgagatcggaggaattcga.....tagagctacttcgaaatag 756

Sequence: 1 ttgagatcggaggaattcga.....tagagctacttcgaaatag 756

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

GenEmbl:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	403.6	53.4	15820	1	AE000737	AE000737 Aquifex a
2	63.8	8.4	11357	1	AE012939	AE012939 Chlorobiu
3	57.2	7.6	11039	1	AE004756	AE004756 Pseudomon
4	55.4	7.3	11618	1	AE013136	AE013136 Thermocana
5	55.2	7.3	2797	1	ECJ224437	AE224437 Erwinia c
6	55	7.3	329100	1	SME591787	AL591787 Sinorhizo
7	53.6	7.1	340806	1	NAI122491	A96528 Sequence 56
8	53.6	7.1	340806	1	NAI122491	AL162752 Neisseria
9	52	6.9	903	6	A96526	A96526 Sequence 55
10	52	6.9	9831	1	AE002562	AE002562 Neisseria
11	52	6.9	172325	6	AX044035	AX044035 Sequence
12	52	6.9	300950	1	AP001516	AP001516 Sequence
13	49.8	6.6	10410	1	AP005340	AP005340 Escherich
14	49.8	6.6	291136	1	AP002556	AP002556 Escherich
15	48.2	6.4	3653	1	ECOTGSA	M64675 Escherichia
16	48.2	6.4	10648	1	AE000221	AE000221 Escherich
17	48.2	6.4	13860	1	D90758	D90758 Escherichia
18	48.2	6.4	16397	1	D90758	D90758 Escherichia
19	48.2	6.4	23203	1	D90852	D90852 E.coli geno
20	48	6.3	4406	1	EC4110309	MJ410309 Erwinia c
21	47.4	6.3	903	6	A96530	A96530 Sequence 56
22	47.2	6.2	2266	1	SFTYRT	X66849 S.flexneri
23	46.6	6.2	22388	1	AE008778	AE008778 Salmonell
24	46.6	6.2	254050	1	AL627269	AL627269 Salmonell
25	46.2	6.1	40138	6	AR118084	AR118084 Sequence
26	46.2	6.1	41587	1	AB025342	AB025342 Moricella
27	46	6.1	998	9	HS435659	MJ325659 Homo sapi
28	45.8	6.1	467	6	AX438333	AX438333 Sequence
29	43.6	5.8	218470	1	BSUB0013	D84432 Bacillus su
30	43.6	5.8	282700	1	BACJH642	AP002936 Mesorhizo
31	43.6	5.8	349619	1	AP002936	AE008096 Agrobacte
32	42.8	5.7	10741	1	AE008096	AE009131 Agrobacte
33	42.8	5.7	11010	1	AE009131	AL590445 chromosom
34	42.8	5.7	211018	3	CNSO7BGC	A96524 Sequence 55
35	42.6	5.6	447	6	A96524	AE003860 Xylella f
36	42.6	5.6	10593	1	AE003860	AE005531 Brucella
37	40.8	5.4	10029	1	AE005531	AE005531 Brucella
38	40.4	5.3	7218	6	IC6494	IC6494 Sequence 14
39	40.2	5.3	253946	2	AC105161	AC105161 Mus muscu
40	39.8	5.3	159505	2	AC023142	AC023142 Homo sapi
41	39.8	5.3	310001	9	AF178030	AF178030 Homo sapi
42	39.6	5.2	5137	1	AF404759	AF404759 Bacteroid
43	39.4	5.2	269223	6	AX067466	AX067466 Sequence
44	39	5.2	9511	1	AF026065	AF026065 Ralstonia
45	39	5.2	12017	6	E22111	B22111 Aromatic co

# ALIGNMENTS

RESULT 1  
AE000737  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AE000737 15820 bp DNA linear BCT 25-MAR-1998  
Aquifex aeolicus section 69 of 109 of the complete genome.  
AE000737.1 GI:2983782  
Aquifex aeolicus.  
Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
Aquifex aeolicus  
1 (bases 1 to 15820)  
Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,  
Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.  
The complete genome of the hyperthermophilic bacterium Aquifex

aeolicus  
JOURNAL Nature 392 (6674), 353-358 (1998)  
MEDLINE 98196666  
PUBMED 9537320  
REFERENCE 2 (bases 1 to 15820)  
AUTHORS Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J. and Swanson, R.V.  
Direct Submission  
Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121

Putative indicates no similarity to known proteins  
Hypothetical indicates similarity to a protein of unknown function.  
FEATURES  
Location/Qualifiers  
1. 15820  
/organism="Aquifex aeolicus"  
/strain="VF5"  
/db\_xref="taxon:63363"  
complement (12..1298)  
/gene="gpsi"  
/note="aq 1367"  
complement (12..1298)  
/gene="gpsi"  
/codon\_start=1  
/transl\_table=11  
/product="glucose-1-phosphate thymidyltransferase"  
/protein\_id="AAC07344.1"  
/db\_xref="GI:2983786"  
/translations="MVEATVLAGGNGRLKPTTEVPKALLKVGARELLYRTIKQLQ  
DVGKFNVI VNNKKFGKVKAFLEKNFEAEVFPNEHPKENGYSLYLAKRIKGEFA  
VNSDHIKFALEKAVEGKGLIVDRGLYINKNEATKVKCEGRKIKYIGKLEKYDG  
FDTGFFVLDESIFEVAEALKEQKLTMSLAKRAOI PCTEVSQVPMVDVTPDEVEK  
AKYLVNTAKTGVGDFISNLARKVSTRISPLVDKFTPNQLTVTLFDLGMFSAIVA  
YESPALGILLQINSLMDGLDGEVARAQKRTKFGAWLSDLDVRYDFAFLSALAWHL  
KPSWFMFWAALFQVMSVSYSTERYKAGYCEDAYAVIKELRYLLGKREDERIFIMI  
FTILGWIKLFFVLAITINLRVLTITLYLVNKKGVN"  
/gene="tagD2"  
/note="aq 1368"  
complement (1313..1819)  
/gene="tagD2"  
/note="aq 1369"  
complement (1313..1819)  
/gene="tagD2"  
/codon\_start=1  
/transl\_table=11  
/product="glycerol-3-phosphate cytidyltransferase"  
/protein\_id="AAC07343.1"  
/db\_xref="GI:2983785"  
/translations="MNRKGVITYTDFLHGHNLKRAKAGDFLIVGVSTDEF  
NAIKKKSVYPEHRAEIVRSIKYVDLVIPERNWEQIEDIKYNDVFMGDDWKGK  
FDYLKEYCEVYVLPRTGISTTELKEALVKLSNILEEDILRLVDIWRGRICIEBILK  
NKLKNGE"  
1794..2576  
/gene="aq 1369"  
1794..2576  
/gene="aq 1369"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07356.1"  
/db\_xref="GI:2983798"  
/translations="MTTFPFLFIVLPFYLMPFVKIFLKLGLGDDMEIQTLVRSKKL  
TPNREFLKNPEYKYLKIENWELQGFQVQKLGNEYLVIYIRPEKEPELGIF  
QSVREAMGVLSTALEHGWEEVPKNYVIHADPVEGNGKLAIAIKTEGISTYDQKL  
EEMKKMVRPRVYVSSDVLTKIDIPDVQSKAVIAREIAKETGSAPELEELGKI  
YGVDTSTLGGLELIEKLLQNPVKLPGGKVNKLKPYWPLEA"  
2573..3088  
/gene="ripA1"  
/note="aq 1370"  
2573..3088  
/gene="ripA1"  
/codon\_start=1  
/transl\_table=11  
/product="rare lipoprotein A"  
/protein\_id="AAC07342.1"

gene  
/db\_xref="GI:2983784"  
/translation="MIALIFPLILSFVSVKELFLEDPLKEPSEPKLFLKEREYKRIF  
RIFVKRDCRYEVGYASVYGPKHYRTANGERNFMFKLTAASTFSLGTYLVYNLE  
NGKFTVIRINDRGFYVDGRIIILSMAAADIEMLKKGKGVAKVLVPLKCLAPSTQVKIY  
BEVVKDIMKTY"  
3099..4472  
/gene="argH"  
/note="aq 1372"  
3099..4472  
/gene="argH"  
/codon\_start=1  
/transl\_table=11  
/product="argininosuccinate lyase"  
/protein\_id="AAC07341.1"  
/db\_xref="GI:2983783"  
/translation="MERFWSGRPKETDKFVEDFTESVDFDKELAFEDIEODIAHVKT  
LOKAGILTEERARELIQBLKIKKEIKEGFOWKKELEDVHMNIEAELINRLGVDGRK  
LHTARNDQVATDEKLYLKEIKVLOLLKELKRTVELAEATTVDVFMPSYTHLQRA  
QPIRVAYFLAYREILLKDSERLMDTRRVDDELPLSGAVAGVDFPLDRFTYALLGF  
NRVTRNSMTAERDFFIIFELNSCALIAQHLRLAEDLIWNTSEEFNVELPDKLCTG  
SSIMPOKNPDVLEIRKGTIRYGNLIALLITMKALPMAYNEDMDKEPFLDPLTKN  
LKNMIGMTLVLSLRVKEONMRKASGNFLITDIANYLVKEGVPPRTAHIAIGSIVA  
YLLEKGGKLEEMTLEEFQSEKPEEDVDFILSPERAADRKRVRVGGTAKKEILLRIEV  
AKAEEG"  
4548..5327  
/gene="nuoH2"  
/note="aq 1373"  
4548..5327  
/gene="nuoH2"  
/codon\_start=1  
/transl\_table=11  
/product="NADH dehydrogenase I chain H"  
/protein\_id="AAC07347.1"  
/db\_xref="GI:2983789"  
/translations="MHSFWINLLFAVIKIVILLIAGIGAFLTWVERKVAAHIOREP  
GPMVWGHLLOPLANGLKLLTKEDLPRYGDPLFHLHAILALYPATILVSVVPFGP  
EFEVFGYKPFILSDVNVGLLLPALGSMAYVALGASNSKIPILIASMRKAGVLV  
SIEVITFAVGMPIMLAGTLYEIVQKIEQNLWYIWQPIAFVVTVMFAALAEHMERV  
PFDVQEAELVTGFTVEYGMKFGLPFLVEWYIEVLPCLL"  
5294..5596  
/gene="nuoH3"  
/note="aq 1374"  
5294..5596  
/gene="nuoH3"  
/codon\_start=1  
/transl\_table=11  
/product="NADH dehydrogenase I chain H"  
/protein\_id="AAC07348.1"  
/db\_xref="GI:2983790"  
/translations="MVRSLTSAIAVVLFPFGWSPINIPFVGFDVDFLPLGLSPVP  
WFILKTLALFLFILMLHWTLPYRIQITQNAWKIMLPLTLFNLVLAFLAPIWR"  
5603..6229  
/gene="nuoI2"  
/note="aq 1375"  
5603..6229  
/gene="nuoI2"  
/codon\_start=1  
/transl\_table=11  
/product="NADH dehydrogenase I chain I"  
/protein\_id="AAC07349.1"  
/db\_xref="GI:2983791"  
/translations="MIKKVAAKPLSWLERIFFIDFKGLRITLKNALRKTITTHYPYS  
KITPPKRFGTFAHKVVDGTEPQPAFAQEWNRVNIILVEYKSCVVCVLCRACVPVQ  
LFPIEGKKLPNGKRVSVFMNMLLCTYCGFCVDACPDCLYQTDIHNENASYTKDAV  
LTLEILEIQIGRDWRREPRPDRIWIDDEQRMKLGNNVVKLPKBEV"  
6226..6762  
/gene="nuoJ2"  
/note="aq 1377"  
6226..6762  
/gene="nuoJ2"  
/codon\_start=1  
/transl\_table=11  
/product="NADH dehydrogenase I chain J"

Query Match	53.4%; Score 403.6; DB 1; Length 15820;
Beet Local Similarity 71.7%; Pred. No. 4.8e-110;	
Matches 529; Conservative 0; Mismatches 209; Indels 0; Gaps 0;	
QY 1 TTGAGATTGAGGAATTGGAAGAAGATTAACCTCGTTCTTCGCGAGAGCTCGAAGAGCG 60	/protein_id="AAC07350.1" /db_xref="rGI:2983792" /translation="MMGRLLTIFSEFSLIALISGIGITLTRPVYVVALISLIVAG IFFTGALVGLAQQLIVAVAVFTIIVISAPMKAKKSSHSRPEELISLPVLV LYLEMTIVPLLVGVKASPEKIAKIFKPFQNTVEVIAKSVKFLAFEVVISVLMGM GAVLIGKESQTEYEDDTA" 6743..7051 /gene="nuoK2" /note="seq.1378" 6743..7051 /gene="nuoK2" /codon_start=1 /transl_table=11 /product="NADH dehydrogenase I chain K" /protein_id="AAC07351.1" /db_xref="rGI:2983793" /translation="MKTIPLDEAPLVTSMLFGLGLIGIITARNNLVTMSLELAINAV NIALVGDHYLGLAEQIFALFTIILAATAEAAVGLDIIAIFRLKVESTDEIRLGG " 7054..8988 /gene="nuoL3" /note="seq.1379" 7054..8988 /gene="nuoL3" /codon_start=1 /transl_table=11 /product="NADH dehydrogenase I chain L" /protein_id="AAC07352.1" /db_xref="rGI:2983794" /translation="MRGLPVIFPPLAFLIILAFGRKIGDLSGIIASVAGLTLFS LVVALKAHSPHVKLIDPLFIGNVTLISGFYFDSLSLALVTFVAVLLFVSIIGY MRBGEKWPVKFAVYSLIFLFLPMLLVISDNMLGIFPMGEVGLASVYLIGYHOKK ATASAFPAFVNRKIDGMLFPGIIVSFYFLKTLIETIIPKREYDKALGATMLLF GAGVSGSGCPPLHTWLPNMAAGTPVSAFLHAATVAAVAVVAVARLYPMFAETPOTLK LVILIGLITWTLAALAGAIVHNDIKKIIFASTMSQLGMEFVALGVGDGMFHLITHA FFKALLFLAAGVITTAFFHHLLNDIPMGGLKCYMPVYVAEMIGLSLIGVPRSGPW SKRIYASMTBMSGVLCVUGCTIYAFITTYARREGRFLVHGRBRREITIKDVHEVEG VMVPEMGLLGFVLVGLFGLMLHMYVGLIGREKGIHLVALVSLGVALGIVLAW AAVVKEVINDKFFKAYESLKEFHTTFKESQFPEKLYHNVIAGVYIVASRVAKVGDITVI DGEINMLYKFFKFEVFLMKCYLDIKIIDLVIHETVLTAFRLGRLSRRLQDTGLVNHYL FLAVGLFELIGLMYLTIDRL"
DB 12343 TTGAAATTGAAGAGGTTGAAGAGGTTAACTTGAAGTTCTTCGCGAGAGGCTCGCAAGGCT 12402	
QY 61 ATAGCCCAATAGGTGTTTGAAGCTATTAACGAGCTCGGTATTAAGGTGAAGGCTTTA 120	
DB 12403 ATGCCCATATAGGTGTTTGAAGCTCTGGAAGAGCTCGGTATTAAGGTGAAGGCTTC 12462	
QY 121 AGCGGGGTGAGCGCGCGGGGCAATCGTTGGTCTTTTATGCGCCAGGCTAATCCCGCTGA 180	
DB 12463 AGCGGGGTGAAGGTGTGAAGCTATGCTTTCGCTTTTATGCGCTTCGCGCTACATCCGAC 12522	
QY 181 GGGATGTTACGCTTTCGAAGAGGGTAACTGGCTGAAGCTGTTTAAGTTCAGCCACCT 240	
DB 12523 GAGATGTTTAAACTCTCGAAGAGGTAACTGGCTCAAACTTTTAAAGTTCAAAACACCG 12583	
QY 241 CTGAAGGATTTGATAGGGTGGGAGAAAGCTATTAAGATTCCTTGAAGAAATTCCTCCTTAC 300	
DB 12583 AAAATGGCTTAAAGGGGTGGGAAAGCTGACAGAGTTTGTGAAGAAAAGCTCGAGATT 12643	
QY 301 AGGAGAAATAGAAAACCTTGAGATCCGACCTATATATGCGCAGCGATTATCTCGGGA 360	
DB 12643 AAGAGGCTCGAAGACCTGAACATACCAACCTATCTTGTCTCGGAGGATCTGTACAGGGA 12702	
QY 361 AGGCTCTATACCTCTCTCGAAGGGAGATTATCCCGGCACTTCTCGGAGCTGTGCAATT 420	
DB 12703 AAGGCTCTTATCTTCTCGGAGGGAGCTTAAATTCCTCGGTGCTCTCGGAAGTTGTTCCATA 12762	

Oy		421	CCCGGCAATTTTGAAACCCGGTGAAGTAATAAGACTCTGCCTGTGACGGAGCATAGTT	480
Dd		12763	CCCGGCAATTTTGAAACCAAGATTGAGTAGAAGAATTTTCATCTGTGTAACGAGGATATAGTG	1282
Oy		481	AACAACCTTCGCCGTGAGGCCCTTTCAGAAAACGGGTATGCCACCCTTGTGCGTGAATGTC	540
Dd		12823	AACAACCTTCGCCGTGAGAACCTTTGGAAAAAGTTCAAAGAACCCATATATCAGGGGTATGATGTG	12888
Oy		541	CTTCCCATTAGAACCGGAAAAAGGATNTTAAAGAACCTTTCAACATCCTTTTGAAGAGCTTC	600
Dd		12943	TTTTCTGCGCGTCTCTTTCCAATTCGAAAAAGGAAAGAGTTCTTCACAGCTGATNTTAA	13000
Oy		661	CCTGAGCTTGAAGAGATTCAACCCCCCTTGANTTTGAAAAACGGAACAATATAGAGAGG	720
Dd		13003	CCTCCCTCTTGAAGAGATTCTCTCCTCTGGAAGTAATTAAGAGGAGCATATATTCCTTGCGG	13060
Oy		721	GGATACATAAAGGCTTAA 738	
Dd		13063	GGATATGAGAGCACCTTTA 13080	
RESULT 2 AE012939/c				
LOCUS	AE012939	11357 bp	DNA	linear BCT 01-JUL-2002
DEFINITION	Chlorobium tepidum TLS	section 160 of 194 of the complete genome.		
ACCESSION	AE012939	AE006470		
VERSION	AE012939.1	GI:21647896		
KEYWORDS				
SOURCE ORGANISM	Chlorobium tepidum TLS. Chlorobium tepidum TLS. Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae; Chlorobium.			
REFERENCE AUTHORS	1 (bases 1 to 11357) Eisen,J.A., Nelson,K.E., Paulsen,I.T., Heidelberg,J.F., Wu,M., Dodson,R.J., Deboy,R., Gwinn,M.L., Nelson,W.C., Haft,D.H., Hickey,E.K., Peterson,J.D., Durkin,A.S., Kolonay,J.L., Yang,F., Holt,I., Umayam,L.A., Mason,T., Brenner,M., Shea,T.P., Parksey,D., Niernan,W.C., Feldblyum,T.V., Hansen,C.L., Craven,M.B., Radune,D., Vamathevan,V., Khouri,H., White,O., Gruber,T.M., Ketchum,K.A., Venter,J.C., Tettelin,H., Bryant,D.A. and Fraser,C.M.	The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)		
TITLE	JOURNAL PUBMED REFERENCE AUTHORS	2 (bases 1 to 11357) Eisen,J.A., Nelson,K.E., Paulsen,I.T., Heidelberg,J.F., Wu,M., Dodson,R.J., Deboy,R., Gwinn,M.L., Nelson,W.C., Haft,D.H., Hickey,E.K., Peterson,J.D., Durkin,A.S., Kolonay,J.L., Yang,F., Holt,I., Umayam,L.A., Mason,T., Brenner,M., Shea,T.P., Parksey,D., Niernan,W.C., Feldblyum,T.V., Hansen,C.L., Craven,M.B., Radune,D., Vamathevan,V., Khouri,H., White,O., Gruber,T.M., Ketchum,K.A., Venter,J.C., Tettelin,H., Bryant,D.A. and Fraser,C.M.	Direct Submission Submitted (30-Apr-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA unpublished	
TITLE JOURNAL REMARK FEATURES source		location/Qualifiers 1..11357 /organism="Chlorobium tepidum TLS" /strain="TLS" /db_xref="taxon:194439" complement(75..3314) /gene="hsdr" /note="CT1878" complement(75..3314) /gene="hsdr" /note="identified by match to TIGR protein family HMW TIGR00348"		

```

/codon_start=1
/transl_table=11
/product="type I restriction system endonuclease"
/protein_id="AAW73097.1"
/db_xref="GI:21647897"
/translation="MWSRWIDYKTIASNNFVILDRYMPDMRWKVAEGYQSEADLERELI
DLRRGGYDPAIKTEPAELANVRVQALNDVQFSDGEWARFVETWLDKPSGIVE
KTRKHIDDYIHDFDQDRIQNIHLVKKTKLVNRKQVIRVFTVTPALCATPREGNF
S3GEQLWDPQPGFVGVKRGVYDVTILVNGLPVQVELKRGVAIRAFANQVHR
YKSEFNSAHLFKYLQLFVINGTDTVFANTTRDKNSDFPMHMAKADNTRDL
KOPAAFPQKHLLSVLHYSVDFVSNLTLLVWRPQIAATERILWKIKSSHQAKTWT
PEGGVIHWTSGKLTSLFKARLGTELDIDKVFVVDKDLDTQMK3YQRPSPD
SVNGSDTAGLRNLEKDNRIIVITIQLNLMKSEPDLPYHKQVVFIDECHRSQ
F3EAQLKKFKYLFQFTGTPTIPPENALGAETASVFGRELSHYITDAIRDEKY
LKFQVNDVRFKAIQFETDERDEKLSAENRQALLHPERIREITQYILTHYOKTHR
LOPAGFNALFAVSVAAKLYEAFKQOKDSAKPLKATIFSYAANEQAQDVGDI
ADGPFVSNLSAKEFNAAIADYNALFKTNFSDVSQGFQNYRDLKRVKGTDDSG
KLPADEQVLLIVGMFUTGDFDAPTLNLFVDKLNRLHGLLQASRNRIFDAPTKF
GNIVTRDLEQATIDAITLFGDKNTRNVVLEKSYREYMEGYDALTGOARRGFVEVQ
ELQARPPDPALEKEADKAFVRLFCGYLRAENVLQNFDEFAALKALQSVNTGPAAV
EAFKQVLLKQSEINLDYILBELIPERNKETSKAELVEEVRVIRASLGHAKESLV
VDFINQTDLEQADKASVLEAFETPARAELQREAEQELIEAKLNAEAARYIATSLK
ERASUTGDLNAVLPMSFLNFQYLTQKQSVFQKIAAFVEKFKGVGGV"
/genes="CT1879"
/complement(3322..3903)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73098.1"
/db_xref="GI:21647898"
/translation="MKETQYKSLPYNPALDRAKALKAGILHEALLWFLKSNKLN
GLDFDEQKLIQNYIVDFYCAERSVIEVDGSHDSKQIEDREDRDLNGLGTVIRVL
ADVLNRLGQVFEPLKDHPLALGTTPPEGKNTALGTTPPEGKNTALGTTPPEGKNT
ALTGTPPEGKNTALGTTPPEGKNTLTLEARE"
/complement(3896..5320)
/notes="hads-2"
/genes="CT1880"
/complement(3896..5320)
/notes="hads-2"
/genes="hads-2"
/notes="identified by match to PFAM protein family HMM
PF01420"
/codon_start=1
/transl_table=11
/product="type I restriction system specificity protein"
/protein_id="AAW73099.1"
/db_xref="GI:21647899"
/translation="MSGREFLQKLQDGERVEWKAIGEIIQLEKGLQNLKLLSSGRY
PAYNGMSYSGFTDSYNSKNTIISQGGASAGFVNTTKFYANAHCVVLPDTEVV
DNRYIYHFLKLAERLTSCQHGAGIPALRASEITSIKIPICPDNPKEKSLAOAIVR
ILDAPTELTAEITARKOVAYVRDLRLTFTTPPYGHPKSGGELSLRHPSEGG
ELPTPYGHSVEERELNSPSLKGWQAPQGVPEWKTGLEVGHFIRSGIOKSPKAS
GVCYIHYGHYGTWTETKSPIDPEANRLKAKPGDLVIATISDEDDAVAKAVA
WGTEDVAVSTDAIYRTPANPKMSIFFQTMDFQEQKPYITGKVRIRSGDNLAKI
LIPILPAASQRIVALDQFDALNTSLTEGLPRELRLQKQYAYVRDLLFSPKASF
GVEPGRDQPSFGVPEGRGLNA"
/complement(5487..7043)
/notes="hads-2"
/genes="hads-2"
/complement(5487..7043)
/notes="hads-2"
/genes="CT1881"
/complement(5487..7043)
/notes="identified by match to PFAM protein family HMM
PF02384"
/codon_start=1
/transl_table=11
/product="type I restriction system adenine methylase"
/protein_id="AAW73100.1"
/db_xref="GI:21647900"
/translation="MTSIQORAEQLQRRIWQIANDVRGTGDMDFQYVYLGALFYRFPIS
ENFAHMEAGDDGIRYAEIPDSVITPELNDDAIKTKGYFIYPSQLFANVAVANTNDS
LNTDLAAIPTAIESSANGYPSEQIKGLFADFTTNSRLGNTVKDKNQRLAAVLKQVA
ELDGPDDDAHIDLFGDAYEFLISNYAANAGKSGGEFTTPOHVSRLIARLALHGOXSV
NKTYDPACSGSLLOAKKPEERLIEDGFRGQESNHTYTNLARMWFLHNTYDKFN
IQCNITLLEPHFADERPDAIVSNPYSVKMIGSDPTLINDERFAPAGVLAPKSKAD
FATVHALHYSKAGRAAIVCFPGIFRGAAEKIRQYLDNNYIVETVIALAPNLFPG
TTAVLNLVLSKHGPDPTTQFIDASALFKFETNNVLLDEHIEQIMAVFASKEEVPHV
AQSVPLERTAAANNYLSVSVSEARDETRVEDIAQLNABELKTTVARIDELRKQIDAIV
ARTGSEDEDA"
/complement(7106..8236)
/genes="CT1882"
/complement(7106..8236)
/notes="identified by match to PFAM protein family HMM
PF00534"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="AAW73101.1"
/db_xref="GI:21647901"
/translation="MKVALYAGTYVKDKGAVRSIYQLVSSMIKNGHQVWVWTPDFTF
EAKASVPVNLTPSVPIPLPYDKLGFVNAVTERQLEFAPDIWHISTPDIVGRKPLHY
AKKGLPVGSVHTDPSPVLSYVRLGFAEPNAVWFLRKFVNAQCDVTLAPNESVRELT
GKIGIRVELMSRGIDKELDPDSRSKLRAMDAGRTVIIYAGRFVLKIDBVMSL
YQFADGGLDVRVFMVIGSPPEAQMRARMPAVFTGYLTGTLTPEAVASGDLFLFP
STTFAFCNVLTALATGLPAVVDGCGQCELVRSQGFVAKAGDVGDVFCCTKLMQ
DGLFRSMRERGLAFADKSWAAVNGKLIDRYLELIAAKARR"
/complement(8408..8875)
/genes="CT1883"
/complement(8408..8875)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73102.1"
/db_xref="GI:21647902"
/translation="MPIRLRLEPDSHQVYFVISHSPAYTAQDIAAAHVSGNELVK
TVNVSIDGKMAALLHAPRLDFDLRLCELRGSDVTLAEIAPSGIFPECEIGAMPFF
GNLYGMKVYADELDESMDIVFNAGTHRELLRLSWFDYKRLVNPVNMGRASIR"
/genes="CT1884"
/8875..9003
/8875..9003
/genes="CT1884"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW73103.1"
/db_xref="GI:21647903"
/translation="MRPPWPCVEPLMFTTEHKFOSVEPARNEKPARHVMVMDTGRNRP"
/complement(9084..9866)
/genes="CT1885"
/complement(9084..9866)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73104.1"
/db_xref="GI:21647904"
/translation="MSTTENVSBRTGLAFGGVVLGAHIVLKAHETGPRACVSG
TSIGSTIANYAFKSWREIREVALELWSDLSGLTSGYLLSIRKPKIVRAQLGS
RRIEDAPLPAIVATDICTGNEVLRGVDVATVMASSSIPIGIFKPEQEGMILVDVG
LTENVPSPLKMGASRIACIDLFGHRSFRPEHLSDLLINAFYSAMRAISQIQISKA
DLTVAPDLRSLFSLVDMSSAVPEILDITGTYREALPLLESWRDAHR"
/complement(9871..11253)
/genes="alda"
/notes="CT1886"
/complement(9871..11253)
/genes="alda"
/notes="identified by match to PFAM protein family HMM
PF00171"
/codon_start=1

```

```

/transl_table=11
/product="aldehyde dehydrogenase"
/protein_id="AA004756.1"
/db_xref="GI:21647905"
/translation="MTYDASTAGLRRCYDGGGTRPFRMRRAQLGLAPLRREHEI
AAVHADLRKPVATMTLPTALRSEIRIVLRLRMRPRKRVGPIHAYOPARAYER
DLRGVILITGAMNYPLOQCLAPLIGLAGGNVSVKPSMAPATISALIASIDGRVDP
QAVRIVEGGEASARLEHCPDHIFFTSRRRQVMSAAHRLPVTLEGGKSPVI
VLEKDLRLAARRIIMAKPLINAGQTCVADYLLVQEGVPEPLQIMKRALIYQSDP

```

Query Match 8.4%; Score 63.8; DB 1; Length 11357;

Best local Similarity 44.0%; Pred. No. 8.6e-08; Mismatches 318; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

```

QY 15 ATTGAAGAGATTAACCTGCTCTTTCGGAGAGAGCTGCAAGAGCATAGCCACATAGG 74
DB 9846 ATCGAAGACACCGGCTGGCCCTTCGGGGCGGCGGTGCTTGCGCGGCCACATCGG 9787
QY 75 TGTTTGAAGACTTAACAGAGCTGGTATTAAGGTGAGGGCTTTAAGCGGGGTAGCGC 134
DB 9786 CGTGCTCAAGCGAGATGAGAGACCGGCTTCAGGGCAGAGTGCCTGCGGCACACGAT 9727
QY 135 CGGGGCAATCGTTGGGCTTTTAAAGCTCAGGCTACTCCCTGAAGGAGATTGAGCCT 194
DB 9726 CGGTTGTTCAATCGCGCGCATGATGCTTCGCGAAGAGCTGGCGGAGATCAAGCAGT 9667
QY 195 TCTGAAGAGGTAAACTGGCTGAGCTGTTTAAGTTCAAGCCACTGTAAGGAGATTGAT 254
DB 9666 CGGCTCGAGCTGAGATGCTGATCTGATCTGCGGCGCTCAACCTTCGCGTACGAGCTCT 9607
QY 255 AAGGTGGAGAAAGGCTATTAAGATTCCTTGAGAAAGTTCTCCCTTAAGAGAAATGAAA 314
DB 9606 CTGATCCCGCAAGTTGGCAAGATCGTCCGCCCGCAGCTCGGTTCCCGCGCATGAGGA 9547
QY 315 ACTTGAGATACGAGATATATATATATATATATATATATATATATATATATATATAT 374
DB 9546 TGGCGCGCTCTCTCTCGCATGTCGCGCACCGACATCTGCAACCGAAGAGAGTGGTCT 9487
QY 375 CTCGAGAGGAGATTAAATCCCGCACTTCTCGGAGAGCTGAGCAATTCGCGCATATTGA 434
DB 9486 GGGCGAGGCGCAGCTGCGCACCGCGGTATGCGAGCTGCTGATTCGCGGATTTTCAA 9427
QY 435 ACCCGTGAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 494
DB 9426 GCGGCTGAGAGAGGAGAGATGCTGCTGATGCGGCGTGAACGAGAAATGTCGCGT 9367
QY 495 TGAAGCCCTTCAGGAAA--GGGATATTCACCGCTTTCGTTGATGTCCTTCCATAGA 551
DB 9366 TTCTCCGCTCAAGAGATGCGCGCGCGAGATCGGTGTCATCGACTTTTTCGCGCGCA 9307
QY 552 GCCGGAAGAGATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 611
DB 9306 CTCCTTCCCGCGCGCGAGAGATCTGTCGATTTGTCCTTAAGCTTTTAAGAGGCCAT 9247
QY 612 CGGCTCAATCTCGAAGAAAGAGATTTTGTGACCTCGTTATAGTTCTGAGCTTGA 671
DB 9246 GCGCGCATTTGCGAATTCAGATCAGCAAGCGCATGTCGTCATCGCTCCGATCTCTC 9187
QY 672 GGAAGTTCAACCCCTGATGTTAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 731
DB 9186 CGGCTCAATCTCGATATATGTCAGCGCTGCGGAGATTTCGACAGCGGATACAGGGA 9127
QY 732 GGC 734
DB 9126 GGC 9124

```

RESULT 3  
 AE004756 11039 bp DNA linear BCT 30-AUG-2000  
 LOCUS Pseudomonas aeruginosa PA01, section 317 of 529 of the complete  
 DEFINITION genome.  
 ACCESSION AE004756 AE004091

VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

TITLE

JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

FEATURES  
 source

gene  
 CDS

gene  
 CDS

gene

CDS

AE004756.1 GI:9949466  
 Pseudomonas aeruginosa.  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 Pseudomonas.  
 1 (bases 1 to 11039)  
 Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
 Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,  
 Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y.,  
 Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,  
 Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.  
 Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen  
 Nature 406 (6799), 959-964 (2000)  
 10984043  
 2 (bases 1 to 11039)  
 Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
 Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,  
 Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y.,  
 Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
 Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H.,  
 Hancock, R.E.W., Lory, S. and Olson, M.V.  
 Direct Submission  
 Submitted (16-MAY-2000) Department of Medicine and Genetics,  
 University of Washington Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 Location/Qualifiers  
 1..11039  
 /organism="Pseudomonas aeruginosa"  
 /strain="PA01"  
 /db\_xref="taxon:287"  
 90..842  
 /gene="PA3335"  
 90..842  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AA006723.1"  
 /db\_xref="GI:9949467"  
 /translation="MSQNDLITGTRESAFLERFPTITLNPDDVPLLOME  
 SLMTPEAMVDIRCDALVEIGCYDGRALLETALINAVLYADLDORALITRTYIEBG  
 MSRDADTVDDILNHTRRGASVGRALYLPLPULGNFRPRKRLDLSIARSTAAVVS  
 VPEDSABATRVROSYTRCGVGLBHTRDGVTFPGSGDFSRYSRACLRHLLABC  
 GLTVRSASNLPHACVTLPEGADQFGSSAA"  
 934..2100  
 /gene="PA3336"  
 934..2100  
 /gene="PA3336"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable MFS transporter"  
 /protein\_id="AA006724.1"  
 /db\_xref="GI:9949468"  
 /translation="MPLIIVVLCNALPALTSTSTRYMAGLMPALAAEFGVSPALIGVY  
 TTYAGAAVGGPLITLALRPERKNALIGLILFVVGGVIGALAPGAVVAVARLVTA  
 VAAAPFGVALITACAEIVBGNQFGASISVLIGLIVGVIGALPVAATWLEWYGRASF  
 VAAVAVAVIGLVLIQIMPAIPGSAGSGSLREBELKVFRRNHLMTVYATISLLIGATFA  
 GFTYFVILTEVSGFSASTVPLLVVYGLATVGNNTVIGRLADRHRTIATVATLIGLAAI  
 AAMVAFALPGOVPAVVALVIVGLGVSNPNPLATVIRGARVGNMNLVNSVPHACIML  
 GMAAGSWIGSLGIAGRGLOGALMVGAALGVALLLTLPLRLRARPVGALGR"  
 complement(2111..3109)  
 /gene="rfad"  
 /note="rfad"  
 /note="PA3337"  
 complement(2111..3109)  
 /gene="rfad"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ADP-L-glycero-D-mannohexose 6-epimerase"  
 /protein\_id="AA006725.1"

```

/db xref="GI:9949469"
/translation="MPMSIIVTGAAGFIGSNLLOALNRGETDIIIAVDLDTGDEQFRN
LADADLADYDNDLERYAGDFVTRALFHQGCACASLENGRYMMENYRYSCRL
LESSLELGVPLVYASAAVSGSRTREARQYERPLNVGYGSKFLFDORVRALPOAR
SOVUGRYNVPYPREHKGEMASVAYHCYQOLRDRGVELRGEHGPFGHLEDFV
AVEDVARVNLHFDHFQRSIGFNLGSGQARTNEVALVINSVRANAOPLSLQOAV
ESGLGFRPESRLARYQSHTCADLELLREAGYRDDPOSLEEGVAGYCRWLARS"
gene
CDS
complement (3199..3489)
/gene="PA3338"
complement (3199..3489)
/gene="PA3338"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06726.1"
/db xref="GI:9949470"
/translation="MPTAKPEIVITYCTOCMLRAAWLAQELLSTFADDLKGVCLEP
GTGGRFRICDGVQWVERKADGGFPEAKALKQRVRDRIDPQRLDGHNDPSPR"
gene
CDS
3558..5744
/gene="PA3339"
3558..5744
/gene="PA3339"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06727.1"
/db xref="GI:9949471"
/translation="MERLLVLLPLLSALAEARPKIGLVLSGAARGLAHIGVLK
ALDEQIIDAIAAGTSMGAVGGLYAGVTPAEELRIALEMDWQOALSAPPRKDVPP
RKQDDDELDVKQKISFRDXTGLGLPLVIGQNLAMVLESLLVHSDNRDFDLAIP
FRVSDTATGEKVPFRKGLHPQAIRASMSIPFAVPAPVIDGRLLVDGGMVNDIPDV
ARMVDVIVIDVGNPLRKDLSTVLDVMQSTITLMTRKNSLAQLATVQNGDVLIIQ
PPLSGTGTDFGRPOLIDAGYRATTVAARLAELRKPHDLNSEALDARTPNQRPV
IDAIRVNNKSKDEVIIRHYIROPLGLTRDLGRLODMSTLYGLDYFOVOYRVYKEK
KLVNTHVAKGKGDTDFRLGLNLSDDMRGSESTNLGGSYRMGNLRIGASMLTRVO
LGRQELYSFYQPLVGRSRYFVAPFLFHEAQNVDTENDPLLYRLERYGYGLNVG
RQTANNGEIRLGAQYKADVRIGDPSLDPDIDTEGYIELKYSFDTVDVNVFHEGE
EIGLWRRYDQSGSDSYRQWDLRLNKALSFGADTWFGGYYGRTLDADAEVYTSFT
LGGARELSGRDALSQNSYSLGRIVYIRLTERSPFLDPLFLYLGGSIERGRIWND
NEYDSGYINAASLMIGFDTPLGLTFESYGINDENKAFYLNQNF"
gene
CDS
complement (5752..7800)
/gene="PA3340"
complement (5752..7800)
/gene="PA3340"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06728.1"
/db xref="GI:9949472"
/translation="MFGKALRVMLCWSCLLAGQASALGVGDIILHSAINQPLDADIE
LLVDGLGADIEIVRLAGADVFAAGVERLOFLNELRFSVLQGRGNRIHVSSIRPV
QEPYLFVLEVPNGRIVREFTVLDPGLYTPRLPLPAARSGLIEPQRSSTPVPAPRS
AAVVDPEALLEPDEYLARPSDNLWASIGRLGAGNADRAQLMEALYOLNPOAFVAD
RHKALARLRLPAGYOPERGAPCAVKEAAVEVLPPADAADVENAPALVAEQROADA
EAEALARQELSQMDLQRLQALQEQLOQRDHQVLELQQLARRQAVRPAAPPA
AFAFPAQVPTFTDSYWRMIVLVLVGLVGLVLRERRRREPFAVEPRKRVAIN
LPLURAPPAAPAPAKVBEQARPFAAPSPPPPPAPAPAPAPAPAPAPAPAPAPAP
DLYVAGRYQOALLQVLAQEPRLSARMKLLVLAELGDAAGDALAEALDLAGG
NPAIDELRGYPALLQMPATPTAATTQDDWSLDPLAESVLPLOPDATSGADGFGD
LNLDLDLWDGALENPLNDPLPRRAAGKAPAEELPAFESNLHELDPVABEYHEFLD
QPEPATVPPAESASLDRACIDSGDLOQASHILRLVVAHGDPWQKAEARELLALIA"
complement (8077..8511)
/gene="PA3341"
complement (8077..8511)
/gene="PA3341"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG06729.1"
/db xref="GI:9949473"
/translation="MAQTDKHYFTGLIAOTSRAWRAELDRRLSHLGLSQARWLVLHL
ARHDSPTQRELAQSVGEGPTLRLDGLSQGLVRLVAEDREAKHIVLTPRADV"

```

```

LIADIEAIAVSNDVLTGIDSEEQALCQOVLRIILANLENR"
complement (8629..9645)
/gene="PA3342"
complement (8629..9645)
/gene="PA3342"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06730.1"
/db xref="GI:9949474"
/translation="MORLIOANGOPHYGIPPSAPOHINVDYDFRSPMGRRVGALGKW
RRHPOFYGLISPELIGGCALQSLGVLGVFVLYQPSGRMLERRFKPLPLGFTFP
SQTPDDGVCSLRQGRNLENMAHREKLLVELDDGTIDACFSEAFSEFQPMCLC
TPTAVNGVYAAQVAGVRCRGTSDSALGDHDLQALDAFAHHDWSAGMRTPTFWNAC
LSEADGLRVGLNLSGCVNETSFTNCYLDLGBLLKVDVSRFDQDPLRPWTIRSY
DGOVELRFEAHGMHQRINLGLLANSFKQIFGCGFSGVLRPRGRAEVRIERLWGFVEDQ
YAKW"
complement (9796..10965)
/gene="PA3343"
complement (9796..10965)
/gene="PA3343"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06731.1"
/db xref="GI:9949475"
/translation="MCVTQKDRDLSRNLAYQDLNRGFRGLRFLPELERYRYRMYLE
DSFALKRALSVMGLVWLFIAGLIDLLIAGPLWGLVLRVGLLVLLVCGCLIMLR
HIOLMPLSTACVLAVGAAAVVGLAHADPGVPVGLLVSPFAAFLAGLRLSQAL
SCALVILLAYLGFEMWAGTQEPGLNNLLFLLFGLNIGAVCYLLEPKSRHFLISRL
RVMDHDSLTGLHNRRSFNHDLRMRQAEOKYKTLALLCDVDHFKAYNDRYQHGAG
DAVQRTIGAVFEANARPLDMARVLRGGEFALLYGANEHEARLAFALRAQVOALRM
EHEASDTAREVTLISVGSCLWPTPGNDLKLRLYDLIYHADRALYEAFAGRNQVVA"
BASE COUNT 1820 a 3727 c 3693 g 1799 t
ORIGIN

Query Match 7.6%; Score 57.2; DB 1; Length 11039;
Best Local Similarity 53.0%; Pred. No. 8.3e-06;
Matches 122; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 23 AGATAAACCTGCTCTTCGGGAGGAGCTGCAAGAGGCATAGCCACATAGGTGCTTTGA 82
|||||
DB 3628 AGATCGGTCTGCTCTTCGGCGGTGCGCCCGCGCTGCGCCATATCGCGTGCTCA 3687
|||||

QY 83 AAGCTATAACAGAGCTCGGTATAGGGTGAGGGCTTTAAGCGGGTGAGCGCGCGGCAA 142
|||||
DB 3688 AGGCCCTCGACGAACAGGCGCATCCAGATCGAGGCCATCGCGGCACCGAGCATGGCGCGG 3747
|||||

QY 143 TCGTTTCGCTCTTTTATGCTCAGGCTACTCCCTCGAAGGATGTTTTCAGCTTCTGAAGA 202
|||||
DB 3748 TGGTCGGCGGCTGTACGCTTCCGCTACACCCCGCGCACTGAGCGCATCGCCCTGG 3807
|||||

QY 203 GGGTAAACTGCTCAAGCTGTTTAAAGTTCAGGCCACCTCTGAAGGGATTG 252
|||||
DB 3808 AGATGACTGCGACAGCGCTGTCCGACGGCGCGCGGCGGACGACGTG 3857
|||||

RESULT 4
AE013136/1
LOCUS 11618 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain MBA4, section 163 of 244 of
the complete genome.
ACCESSION AE013136
VERSION AE013136.1
KEYWORDS GI:20516851
SOURCE Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis
Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
Thermoanaerobacteriaceae; Thermoanaerobacter.
REFERENCE
1 (bases 1 to 11618)
Bao,Q., Tian,Y., Li,W., Xu,Z., Xuan,Z., Hu,S., Dong,W., Yang,J.,
Chen,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y., Ling,L.,
Tan,H., Chen,R., Wang,J., Yu,J. and Yang,H.

```

TITLE A Complete Sequence of the *T. tengcongensis* Genome  
JOURNAL Genome Res. 12 (5), 689-700 (2002)  
MEDLINE 21992816  
PUBMED 11997336  
REFERENCE 2 (bases 1 to 11618)  
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China  
REFERENCE 3 (bases 1 to 11618)  
AUTHORS Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China  
REFERENCE 4 (bases 1 to 11618)  
AUTHORS Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China  
FEATURES  
Source Location/Qualifiers  
1. .11618  
/organism="Thermocaneerobacter tengcongensis"  
/strain="WB47"  
/db\_xref="taxon:119072"  
/complement(94. .277)  
/gene="TTE1837"  
/complement(94. .264)  
/gene="TTE1837"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA025028.1"  
/db\_xref="GI:20516852"  
/translation="MECPNCKSTNKGKGNLTYCRDNCCEIKKKCTAVVMYDAEG  
CVRKPRVCYNA"  
/complement(273. .277)  
/gene="TTE1837"  
/complement(622. .1485)  
/gene="RsaA"  
/note="TTE1838"  
/complement(622. .1485)  
/gene="RsaA"  
/note="Best Blastp hit = gi|1664803|sp|O34731|YLBK\_BACSU  
HYPOTHEICAL 28.3 KDA PROTEIN IN CTAG-RPMF INTERGENIC  
REGION gi|7474645|pir||H69874 conserved hypothetical  
protein ylbk - Bacillus subtilis  
gi|2340007|emb|CAB1357.1| (Z98682) ylbk protein [Bacillus  
subtilis] gi|2633875|emb|CAB1377.1| (Z99111) similar to  
hypothetical proteins [Bacillus subtilis], score 155,  
E-value 6.00E-37"  
/codon\_start=1  
/transl\_table=11  
/product="predicted esterase of the alpha-beta hydrolase  
superfamily"  
/protein\_id="AA025029.1"  
/db\_xref="GI:20516853"  
/translation="MRPRVGLIGGAGARGVHLGLIKPEENIDIPDIVIGSMGAI  
IGAIYAGHVVDKLISDAKKINMLKPSLIDFASRGTGVKGRKIKYGLVGYKESPE  
ELNPLIVATIDTQGTGKGVKEDGLIKARISIDPAFEPVRYGTGLVDGSIYDS  
EATLALASLGADIIINCIVSSIDMGFFERFPLNSKEVLIKKYFNIRKRLPPII  
SITTTAKILKDKNSKMLEKMGKNKYVTIKKPVNDIRWRFPQAKCINMGREADS  
IVARIKRELEK"  
/complement(637. .1485)  
/gene="RsaA"  
/note="Pfam match to entry UPF0028, Uncharacterized  
protein family UPF0028, score 39.6, E-value 2.80E-12"  
/complement(1501. .2084)  
/gene="AcrR3"  
/note="TTE1839"  
/complement(1501. .2070)  
/gene="AcrR3"

misc\_feature  
/note="Best Blastp hit = gi|481591|pir||S38906  
hypothetical protein 4 - Clostridium pasteurianum  
gi|431950|emb|CAA8221.1| (Z28353) similar to a B.subtilis  
gene (GB: BACHEMEHY-5) [Clostridium pasteurianum], score  
64.3, E-value 1.00E-09"  
/codon\_start=1  
/transl\_table=11  
/product="Transcriptional regulator"  
/protein\_id="AA025030.1"  
/db\_xref="GI:20516854"  
/translation="WESKQSOQRIIAAAYCISTKGYANSLDIDIEAAYVLSOLHY  
VPSKRGHFTREVIKMKINKYLKEINELMSGEPKAKQSLVNFPPDLKNNGLPEL  
LYDPTGAWMSSEMSILKDLFNDLSKMEIKLSNLSALGENFRNYSPPAVAMITGA  
MFGQIVLIDNSNEBILDMALAQIIFE"  
/complement(1903. .2043)  
/gene="AcrR3"  
/note="Pfam match to entry tetr, Bacterial regulatory  
proteins, tetr family, score 67.4, E-value 3.00E-16"  
/complement(2080. .2084)  
/gene="AcrR3"  
/complement(2100. .3788)  
/gene="AcrP2"  
/note="TTE1840"  
/complement(2100. .3776)  
/gene="AcrP2"  
/note="Best Blastp hit = gi|10175329|dbj|BAB06427.1|  
(AP001516) ABC transporter [Bacillus halodurans], score  
373, E-value 1.00E-102"  
/codon\_start=1  
/transl\_table=11  
/product="predicted unusual protein kinase"  
/protein\_id="AA025031.1"  
/db\_xref="GI:20516855"  
/translation="MGLRRYRRIIPVITYGRGAIIDNIGILKHINVRKIKOTND  
ENIKLSRGERLRLBELGPTFKMQQIISTSDIIPKQITKLEKLDQKAPAFSD  
EVKSVIONEFSESLAEVAEPPLAAASIAQVHKLMSGKTVVAVKVRPGIKTI  
AQMRLIEDIAKPDVNDHTKGYKINFTKRVEDFQRLDEEDFRIEENAKPEKQNTL  
KDKVKYPSIITHTTRRLVMEYIGGIPLNDPVAIDPAGIDRGALRNALKATLNOI  
LRDGFPGDHPHGNIMVLBDGTLAFDRGWSLSPRKQFSGKLLGIVKNSRMII  
ESIIDNAVITNVNMKLEKDINLRPRVYEIPREKUGSVANGITDIPVSNVIVP  
NEFMALASLTITLIGIVKLDPKISVEVAAPKIKQILPKVFTQHKKESTINATMY  
SRLIKELPSFLNPLRTEENVAIEKIRPLENLEKRVKVFNRSTSLILLALSTV  
IAGILIGSGMSANAGAWYKNGIILIKGIALIARIVYGLAISIFRSGRLL"  
/complement(3784. .3788)  
/gene="AcrP2"  
/complement(3811. .4150)  
/gene="TTE1841"  
/complement(3811. .4137)  
/gene="TTE1841"  
/note="Best Blastp hit = gi|10175330|dbj|BAB06428.1|  
(AP001516) BH2709  
unknown [Bacillus halodurans], score 70.5, E-value  
5.00E-12"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA025032.1"  
/db\_xref="GI:20516856"  
/translation="WTSILERSINFGGLFPLSBEKLEKIVELVDGEVARBDQAOKA  
VMDLVKKEBEKBEIRKMIKDAVAVETGVMVNAKKEIDVTREIKSIYREVRVLE  
KQNKLEK"  
/complement(4146. .4150)  
/gene="TTE1841"  
/complement(4233. .4994)  
/gene="PncA"  
/note="TTE1842"  
/complement(4233. .4979)  
/gene="PncA"  
/note="Best Blastp hit = gi|7517111|pir||C72463  
hypothetical protein APr2350 - Aeropyrum pernix (strain  
KI) gi|5106052|dbj|BA081363.1| (AF000064) 231aa long  
hypothetical protein [Aeropyrum pernix], score 60.5,  
E-value 2.00E-08"

```
/codon_start=1
/transl_table=11
/product="Amidases related to nicotinamidase"
/protein_id="AAM25033.1"
/db_xref="GI:20516857"
/translation="MNFDSFLYNTIRPLNYLFDYFYNLEKSLSSVYINAGBQKVS
IVDMLAGFCCKGLASPRVAGIIEPIKLLKACRYMGIKNVFLNDHPSDAVPEGE
FPCHVKGTPESEIVDELKEIEGEPIVKEKSLNVPFGESEGGNEPLKXVEMIK
KSTFIVVGDCTDLVCYTQAMSLKMANANLKNVIVPENCVEITDTSVKTAQSLKI
NPHDNLHITMFLYHMKLNGIEVVKELLE"
/complement (4990..4994)
/gene="PncA"
/complement (5034..5472)
/gene="Paal2"
/note="TTBL843"
/complement (5034..5459)
/gene="Paal2"
/note="Best Blastp hit = gi|11499845|ref|NP_071089.1|
conserved hypothetical protein [Archaeoglobus fulgidus]
gi|3334444|sp|O28020|YM64_ARCFU_HYPOTHETICAL_PROTEIN
AF2264 gi|7430238|pir|[H69532 conserved hypothetical
protein Ar2264 - Archaeoglobus fulgidus
gi|2648253|gb|AAB8986.1| (AE000948) conserved
hypothetical protein [Archaeoglobus fulgidus], score 87.4,
E-value 7.00E-17"
/codon_start=1
/transl_table=11
/product="uncharacterized protein, possibly involved in
aromatic compounds catabolism"
/protein_id="AAM25034.1"
/db_xref="GI:20516858"
/translation="MKAINSGIDEKLEPEDLLKVNBDTNPHOLIGVHVBLGGYAT
IEIEKHLNPLNIAHGVSFVMDITMGMAARTVGQVITININLYSPVRGKVK
AKGKIHVAGSKTIVACEAYAEIDGRLIAARETFNVL"
/complement (5043..5405)
/gene="Paal2"
/note="Pfam match to entry DUF157, Uncharacterized protein
Faai, COG2050, score 103.2, E-value 5.10E-27"
/complement (5468..5472)
/gene="Paal2"

misc_feature
7.3%; Score 55.4; DB 1; Length 11618;
Best Local Similarity 46.7%; Pred. No. 2.9e-05;
Matches 216; Conservative 0; Mismatches 241; Indels 6; Gaps 1;

Qy 26 TAAACCTGTTCTTTCGGAGGAGCTCAAGGCGATAGCCACATAGGTGTTTCAAG 85
Db 1472 TAGGACTTATCTGGCGGTGGAGCAGAGGATATGCGCACTTAGGTATATAA 1413
Qy 86 CTATAACGAGCTCGGTATAGGTGAGGCTTTAAGCGGGGTAGCGCGGGCAATCG 145
Db 1412 GATTGAGAGAGAAAATTCCTATGATTTTATAGTAGCATAGTATGGGAGCGATA 1353
Qy 146 TTTCCGTCCTTTATGCTCAGGCTACTCCCTCAAGGATGTTTCAGCCTTCTCAAGAGG 205
Db 1352 TAGGAGCAATCTATGCTTCAGGACATATAGTACAACTTATAGTGATGCAAAAAGA 1293
Qy 206 TAACTGCTGAAGCTGTTTAAAGTTCAGGCCACTCTGAAGGATGATAGGTGGGAGA 265
Db 1292 TTAATATGTTAAATTTATAAGTTTATGATTTTAAAGCTCAGCAACCGGCTAGTAA 1233
Qy 266 AGGCTATAAGATTCCTTGAAGAGTCTCCCTTACAGG-----AGATAGAAACTTG 319
Db 1232 AGGCTGAAAATATAGAGAAATATTTGCGGGGTATGTGAAGAAAGTTTGAAGAACTTA 1173
Qy 320 AGATACCGACATATATGCGCGAGGATTTATCTCGGAAGGGCTCTATACCTCTCG 379
Db 1172 ATATTCCTCTATATTTGTGCTACTATATTCAAACAGCAAGGAGTAGTATTTAAG 1113
Qy 380 AAGGAGTTTAAATCCCGCACTCTCGGAGCTGTGCAATCCCGGCAATTTGAACCCG 439
Db 1112 AGGAGACTTGATAAAAGCTATAAGGGCAAGCAATTTCTATCCCTGCTTTTGAACCTG 1053
```

Qy 440 TTGAGTATAAGATTACTTCTCTGCTGACGAGGTATAGTTAA 482  
Db 1052 TGGAGTACATGAACCAAGTTAGTTGATGATCAATTGTTGA 1010

RESULT 5  
ECJ224437  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

ECJ224437.1 GI:3319926  
Erwinia carotovora subsp. carotovora expL, expM, and galu genes.  
AJ224437  
expL gene; expM gene; galu gene; response regulator.  
Pectobacterium carotovorum subsp. carotovorum.  
Pectobacterium carotovorum subsp. carotovorum  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Pectobacterium.

REFERENCE  
1 (bases 1 to 2797)  
Andersson, R.A., Palva, E.T. and Pirhonen, M.  
The response regulator expM is essential for the virulence of  
Erwinia carotovora subsp. carotovora and acts negatively on the  
sigma factor RpoS (sigma s)  
Mol. Plant Microbe Interact. 12 (7), 575-584 (1999)  
99407916  
PUBMED  
10478478

REFERENCE  
2 (bases 1 to 2797)  
Andersson, R.A.  
Direct Submission  
Submitted (13-MAR-1998) Andersson R.A., Plant Biology, Swedish  
University of Agricultural Sciences, Box 7080, 750 07 Uppsala,  
SWEDEN

FEATURES  
Location/Qualifiers  
1..2797  
/organism="Pectobacterium carotovorum subsp. carotovorum"  
/strain="GCC3193"  
/db\_xref="taxon:555"  
344..1264  
/gene="expL"  
344..347  
/gene="expL"  
353..1264  
/gene="expL"  
/note="rseA homologue"  
/codon\_start=1  
/transl\_table=11  
/product="ExpL protein"  
/protein\_id="CAA11941.1"  
/db\_xref="GI:3319927"  
/db\_xref="SPTREMBL:O86196"  
/translation="MQRKKIGALGSGAAKGAHIGVFNALAEIGDIDVDVAGCSIGA  
LVGAAYATNLSSMEQWVGFYMDVIRLMDLSWQSGLLRGDRFVNSVHLLHTTQI  
EDCAIKYGVVTTNLSTGRELWLTGDLHQAMRASCMPGLSPVRFNDYVLVDGAVVN  
FVPSLARAGADIVIAVDLQHDASLNHQDLISIKPTASEVDIDMEHVSQDWRRIRE  
RLLRGRQSAESSPTAMEINSTSIQILENRLKMTWRAGDDPDVLLQFVPCQIATLDFH  
RAQRAIRAGYKAVAKMRDELLPLATEL"  
1370..2397  
/gene="expM"  
1370..1373  
/gene="expM"  
1381..2397  
/gene="expM"  
/function="regulation of virulence factors"  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="ExpM protein"  
/protein\_id="CAA11942.1"  
/db\_xref="GI:3319928"  
/db\_xref="SPTREMBL:O86197"  
/translation="MEQFLAGKHILVIDEAVFRSVLAGYLTSLGASVREAINGLDAL  
SILHYPDLMICDLKMPMMGGIEFLERLKDSDTFLVLSATSMADIAKVLRLGV  
QDVLKPIRDLRLDVMSCLYPDMETSQLNDELQDMFCLNQLNSQSPYATIKLLAQ  
QPVQOQTARCRVNYRQLTATBQPLVLDIAALSETELAFVCLDVTGVNNGTAAAL  
LLRTLNLGLQELHVDQQRLLPYLPTLLKQVNLRLQASLDGRFLLVGYVYHRLKQL



IIISAGIANTLVNVEQVVLNSGVPLGTILEGATVLAQVLYOCETWOCQIWGQGRILRLM  
LTTRE"  
gene 2593. .2797  
/gene="galu"  
RBS 2593. .2597  
/gene="galu"  
CDS 2604. .>2797  
/gene="galu"  
/codon\_start=1  
/transl\_table=11  
/product="GalU protein"  
/protein\_id="CAA11943.1"  
/db\_xref="GI:3319929"  
/db\_xref="SPTREMBL:O86198"  
/translation="MSIVRKVKVKAIVPAGIGTMRLPATKAIPKRMPLVYKPLIYQ  
VVEICIAINIIIVTHSS"

BASE COUNT 703 a 562 c 756 g 776 t  
ORIGIN

Query Match 7.3%; Score 55.2; DB 1; Length 2797;  
Best Local Similarity 54.4%; Pred. No. 3.2e-05;  
Matches 111; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 341 CGACGATTATATCTCGGAGAGGCTCTATCTCTCGGAGGAGTTATCCCGCAC 400  
DB 690 CGACCAATCTGACGACGGGGCGTGAACCTGCTGACAGAGCCGATTTACACAGGCA 749  
QY 401 TTCTCGGAGCTGTGCATTTCCCGGCAATTTGAACCCGTTGAGTATAGATTAATCTTGC 460  
DB 750 TCGGGGCGCTTTCGAGTATGCTGCTGCTTTATCCCGCTTGATTAACATTAATCTGAC 809  
QY 461 TCGTTGACGAGAGTATGTTAAACCTTCCCGTTAGACCTTTGAGAAACCGGATATTC 520  
DB 810 TGGTGGACGGCGCGGGGTGTTATCTCTCCGCTTCTGCGCGGCGCAATGGGGGCTG 869  
QY 521 CCACGGTTGGCTGATGTCCTTC 544  
DB 870 ATATGCTGATGCTGTTATCTTC 893

RESULT 6  
SME591787/c 329100 bp DNA linear BCT 05-JUL-2002  
LOCUS SME591787 329100 bp DNA linear BCT 05-JUL-2002  
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 6/12.  
ACCESSION AL591787 AL591688  
VERSION AL591787.1 GI:15074266  
KEYWORDS  
SOURCE  
ORGANISM Sinorhizobium meliloti.  
Sinorhizobium meliloti  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Sinorhizobium.  
1 (bases 1 to 329100)  
Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J.,  
Boisnard, P., Becker, A., Boutry, M., Cadieu, B., Dreano, S., Gloux, S.,  
Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Maury, D.,  
Pohl, T., Portetelle, D., Puehler, A., Purnelle, B., Rameperger, U.,  
Renard, C., Thebault, P., Vandenbol, M., Weidner, S. and Galibert, F.  
Analysis of the chromosome sequence of the legume symbiont  
Sinorhizobium meliloti strain 1021  
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)  
2136507  
MEDLINE 11481430  
PUBMED 2 (bases 1 to 329100)  
REFERENCE  
AUTHORS Gouzy, J.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELIIO  
EU Consortium  
MELIIO EU Consortium:  
Laboratoire de Biologie Molculaire des Relations  
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,  
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,  
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,  
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,

FEATURES  
source  
gene  
tRNA  
CDS  
gene  
CDS  
gene  
CDS

1. .329100  
/organism="Sinorhizobium meliloti"  
/strain="1021"  
/db\_xref="taxon:382"  
152. .227  
/gene="tRNA-Thr\_TGT"  
152. .227  
/gene="tRNA-Thr\_TGT"  
/product="tRNA-Thr"  
/note="codon recognized: ACA; predicted by tRNAscan-SE"  
/evidence="not experimental"  
complement (373. .798)  
/gene="SMC04435"  
complement (373. .798)  
/gene="SMC04435"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical"  
Gene name confidence : hypothetical  
predicted by Homology"  
/codon\_start=1  
/evidence="not experimental"  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAC45913.1"  
/db\_xref="GI:15074267"  
/db\_xref="SPTREMBL:Q92019"  
/translation="MKNALVEFPAVDWQASKEFCNIIIGFSCDYERPEEGCYSLD  
GALMTIDQIGTGTADGHLPRSPYRGGLNVOIRVASDALVRLTDRGIALYLPLE  
ERWYRRGDEVGRCFVADPDGYLRFYEPGRPRRTS"  
complement (914. .1855)  
/gene="SMC01330"  
complement (914. .1855)  
/gene="SMC01330"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : putative"  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence="not experimental"  
/transl\_table=11  
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAC45914.1"  
/db\_xref="GI:15074268"  
/db\_xref="SPTREMBL:Q92018"  
/translation="MODLNDIALPAVVGNGFSAARLNVKSKLAKHVARLEBOL  
GYRLERSRTRKRMTEVGRIFVEHOGGLDGVAAAEATIAVARPTGYRACIGF  
TMDLADITLAFRRRVGYVLLTASRRRIDLBRIIDVALRARDLDTDSQITKIFG  
EVRQRLASPTLARIGETTNTLSQMPFLSNBOHPNVWRLVHATGDALEIARPV  
VGSQDFLIERAIBGMGIALPDHICBAFPTGALVPTLPMTSGNVHVLVPSRH  
GLLPATRALIDFLAENTLIKALERCHEVPPRPAASFEI"  
2063. .2683  
/gene="acpD OR SMC01329"  
2063. .2683  
/gene="acpD OR SMC01329"  
/function="small molecule metabolism; fatty acid  
biosynthesis"  
/note="Product confidence : probable"  
Gene name confidence : putative  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence="not experimental"

```

/transl_table=11
/product="PROBABLE ACYL CARRIER PROTEIN PHOSPHODIESTERASE"
/protein_id="CAC45915.1"
/db_xref="GI:15074269"
/db_xref="SPTREMBL:Q92017"
/translation="MKILHDSGLIGSHSVSRRLTAISVLSQKADRPDAEITYRDLAS
ERVPHLTCAQIMAPADLEGVLDALLADVRIGROMLREFLAADVTVVVGAPWYFNSIPSQ
LKAWIDRLAVAGKTRFYTEAGPEGLAKGKLLIVASTRGHYSVAPASAMDHQETYLRS
VEGPPGTDIEFTIRAGLNLPQKQPAIAEABKTAEGNVLKLAS"
2792..3076
/gene="SMC01328"
CDS
2792..3076
/function="miscellaneous; hypothetical/partial homology"
/translation="MAKSFERNKPHVNIIGTIGHVDHGKTSLTAAITKYFGFKAQDQ
IDAAPEKAGITITSTARVEYETNRHYAVDCFGHADYVKNMTTGAQMDGALLVVS
AADGPMQTRREHILLARQVPAIWFVLFNKVDQVDDAELELLELVRELLSYEPFG
DDIPVKSALAEALDSKKIGEDAIRELMAADVAYITPERPDPQPLMEIPEVFSI
SGRTVTVGRVERGIKVKGSEIEIVGIRPTTKTCTGVEMPRKLLDQAGADNIGALL
RGVDRNGVERQIILCKFGSVKPRKFAEAYILTKESGGRRHTPFTNRPQFYPRITD
VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGRTVVGAGIVASIV"
6125..6790
/gene="SMC01325"
CDS
6125..6790

/transl_table=11
/product="PROBABLE ELONGATION FACTOR TU PROTEIN"
/protein_id="CAC45918.1"
/db_xref="GI:15074272"
/db_xref="SPTREMBL:Q925Y6"
/translation="MAKSFERNKPHVNIIGTIGHVDHGKTSLTAAITKYFGFKAQDQ
IDAAPEKAGITITSTARVEYETNRHYAVDCFGHADYVKNMTTGAQMDGALLVVS
AADGPMQTRREHILLARQVPAIWFVLFNKVDQVDDAELELLELVRELLSYEPFG
DDIPVKSALAEALDSKKIGEDAIRELMAADVAYITPERPDPQPLMEIPEVFSI
SGRTVTVGRVERGIKVKGSEIEIVGIRPTTKTCTGVEMPRKLLDQAGADNIGALL
RGVDRNGVERQIILCKFGSVKPRKFAEAYILTKESGGRRHTPFTNRPQFYPRITD
VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGRTVVGAGIVASIV"
6125..6790
/gene="SMC01325"
CDS
6125..6790

Query Match 7.3%; Score 55; DB 1; Length 329100;
Best Local Similarity 46.0%; Pred. No. 4.4e-05;
Matches 227; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 24 GATAAACCTCGTCTTTTCGGGAGGAGCTGCAAGGGGATACGCCACATAGGTGTTTGA 83
DB 78470 GATTGCGTGGCGTTCGGCGCGGTGCTCGCGGGGCTGGGCCCATATCGGTGCTCTGA 78411

QY 84 AGCTATAAACGAGCTCGGTATAGGGTCAAGGGCTTTAAAGCGGGTGAGCGCGGGCAAT 143
DB 78410 AGCCCTCGACGAGGAGGATATCGAAGTCGGAATGATTGCGGGCACATCGATCGGTGCGCT 78351

QY 144 CGTTCGCTCTTTTATGCTCAGGCTACTCCCTCGAAGGGAGTGT-----CAGCCTTCT 197
DB 78350 GGTGGCGGCTGTTTACCTCGCGGCAAGCTCGATGAATTTGAGGCTTTTCGCCCTCGCT 78291

QY 198 GAAGAGGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACTCTGAAGGATGATAGG 257
DB 78290 TACCGTGGCGCGATTGCGCGCTTCTCGATTTCGCGATCGCGCGGTGCGCTTCTCGG 78231

QY 258 GTGGAGAGAGCTATAAGATTCTTTGAGGAAGTTCTCCCTTTACAGGAGAAATAGAAAACT 317
DB 78230 CGGTCTCGCGCTCACCAGCGGATGCGAGAACATCTGCAGAACCTCAGTATCGAGGATCT 78171

QY 318 TCAGATACCGAGCTATATATGCGGCGAGGATTTATCTCGGGAAGGCTCTATACCTCTC 377
DB 78170 CGACCGGCGCTTCTGTCGCGCTGCGCCACGAGGCTCTACAGCGGTACAGAGGTATGATCGA 78111

QY 378 GGAAGGAGATTTAATCCCGCAGCTTCTCGGCGAGCTGTCGAATTCGCCGATATTTGAACC 437
DB 78110 GAAAGGATCCCTTATACCGCCATTTCGCGATCTTACGCTTGCCTGGCATTTTCGAGCC 78051

QY 438 CGTTGAGTATAAGAATTACTTGTCTGTTGACGGAGGTATAGTTAAACACCTTCCCGTTGA 497
DB 78050 CGTCAAGCCCAATGGCGGAGCTCTCATCGACGCGCGCTGTGTCAACCCCGTTCGCGTTTC 77991

QY 498 GCCCTTTCAGGA 510
DB 77990 GGTCTGCGCGCA 77978

RESULT 7
A96528
LOCUS A96528 903 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 561 from Patent WO924578.
ACCESSION A96528
VERSION A96528.1 GI:6780156
KEYWORDS
SOURCE unidentified.

```

```

/transl_table=11
/product="PROBABLE ACYL CARRIER PROTEIN PHOSPHODIESTERASE"
/protein_id="CAC45915.1"
/db_xref="GI:15074269"
/db_xref="SPTREMBL:Q92017"
/translation="MKILHDSGLIGSHSVSRRLTAISVLSQKADRPDAEITYRDLAS
ERVPHLTCAQIMAPADLEGVLDALLADVRIGROMLREFLAADVTVVVGAPWYFNSIPSQ
LKAWIDRLAVAGKTRFYTEAGPEGLAKGKLLIVASTRGHYSVAPASAMDHQETYLRS
VEGPPGTDIEFTIRAGLNLPQKQPAIAEABKTAEGNVLKLAS"
2792..3076
/gene="SMC01328"
CDS
2792..3076
/function="miscellaneous; hypothetical/partial homology"
/translation="MAKSFERNKPHVNIIGTIGHVDHGKTSLTAAITKYFGFKAQDQ
IDAAPEKAGITITSTARVEYETNRHYAVDCFGHADYVKNMTTGAQMDGALLVVS
AADGPMQTRREHILLARQVPAIWFVLFNKVDQVDDAELELLELVRELLSYEPFG
DDIPVKSALAEALDSKKIGEDAIRELMAADVAYITPERPDPQPLMEIPEVFSI
SGRTVTVGRVERGIKVKGSEIEIVGIRPTTKTCTGVEMPRKLLDQAGADNIGALL
RGVDRNGVERQIILCKFGSVKPRKFAEAYILTKESGGRRHTPFTNRPQFYPRITD
VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGRTVVGAGIVASIV"
6125..6790
/gene="SMC01325"
CDS
6125..6790

/transl_table=11
/product="PROBABLE ELONGATION FACTOR TU PROTEIN"
/protein_id="CAC45918.1"
/db_xref="GI:15074272"
/db_xref="SPTREMBL:Q925Y6"
/translation="MAKSFERNKPHVNIIGTIGHVDHGKTSLTAAITKYFGFKAQDQ
IDAAPEKAGITITSTARVEYETNRHYAVDCFGHADYVKNMTTGAQMDGALLVVS
AADGPMQTRREHILLARQVPAIWFVLFNKVDQVDDAELELLELVRELLSYEPFG
DDIPVKSALAEALDSKKIGEDAIRELMAADVAYITPERPDPQPLMEIPEVFSI
SGRTVTVGRVERGIKVKGSEIEIVGIRPTTKTCTGVEMPRKLLDQAGADNIGALL
RGVDRNGVERQIILCKFGSVKPRKFAEAYILTKESGGRRHTPFTNRPQFYPRITD
VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGRTVVGAGIVASIV"
6125..6790
/gene="SMC01325"
CDS
6125..6790

Query Match 7.3%; Score 55; DB 1; Length 329100;
Best Local Similarity 46.0%; Pred. No. 4.4e-05;
Matches 227; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 24 GATAAACCTCGTCTTTTCGGGAGGAGCTGCAAGGGGATACGCCACATAGGTGTTTGA 83
DB 78470 GATTGCGTGGCGTTCGGCGCGGTGCTCGCGGGGCTGGGCCCATATCGGTGCTCTGA 78411

QY 84 AGCTATAAACGAGCTCGGTATAGGGTCAAGGGCTTTAAAGCGGGTGAGCGCGGGCAAT 143
DB 78410 AGCCCTCGACGAGGAGGATATCGAAGTCGGAATGATTGCGGGCACATCGATCGGTGCGCT 78351

QY 144 CGTTCGCTCTTTTATGCTCAGGCTACTCCCTCGAAGGGAGTGT-----CAGCCTTCT 197
DB 78350 GGTGGCGGCTGTTTACCTCGCGGCAAGCTCGATGAATTTGAGGCTTTTCGCCCTCGCT 78291

QY 198 GAAGAGGTAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACTCTGAAGGATGATAGG 257
DB 78290 TACCGTGGCGCGATTGCGCGCTTCTCGATTTCGCGATCGCGCGGTGCGCTTCTCGG 78231

QY 258 GTGGAGAGAGCTATAAGATTCTTTGAGGAAGTTCTCCCTTTACAGGAGAAATAGAAAACT 317
DB 78230 CGGTCTCGCGCTCACCAGCGGATGCGAGAACATCTGCAGAACCTCAGTATCGAGGATCT 78171

QY 318 TCAGATACCGAGCTATATATGCGGCGAGGATTTATCTCGGGAAGGCTCTATACCTCTC 377
DB 78170 CGACCGGCGCTTCTGTCGCGCTGCGCCACGAGGCTCTACAGCGGTACAGAGGTATGATCGA 78111

QY 378 GGAAGGAGATTTAATCCCGCAGCTTCTCGGCGAGCTGTCGAATTCGCCGATATTTGAACC 437
DB 78110 GAAAGGATCCCTTATACCGCCATTTCGCGATCTTACGCTTGCCTGGCATTTTCGAGCC 78051

QY 438 CGTTGAGTATAAGAATTACTTGTCTGTTGACGGAGGTATAGTTAAACACCTTCCCGTTGA 497
DB 78050 CGTCAAGCCCAATGGCGGAGCTCTCATCGACGCGCGCTGTGTCAACCCCGTTCGCGTTTC 77991

QY 498 GCCCTTTCAGGA 510
DB 77990 GGTCTGCGCGCA 77978

RESULT 7
A96528
LOCUS A96528 903 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 561 from Patent WO924578.
ACCESSION A96528
VERSION A96528.1 GI:6780156
KEYWORDS
SOURCE unidentified.

```

ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 903)
AUTHORS	Pizza, M., Scariato, V., Rappuoli, R., Grandi, G. and Masiagnani, V.
TITLE	Neisseria antigens
JOURNAL	Patent: WO 92/4578-A 561 20-MAY-1999;
	PIZZA MARIAGRAZIA (IT), SCARIATO VINCENZO (IT), RAPPUOLI RINO (IT),
	CHIRON SPA (IT), GRANDI GUIDO (IT), MASIGNANI VEGA (IT)
FEATURES	
Source	1. 903 Location/Qualifiers
BASE COUNT	205 a 231 c 250 g 210 t 7 others
ORIGIN	
Query Match	7.1%; Score 53.6; DB 6; Length 903;
Best Local Similarity	44.5%; Pred. No. 9.1e-05;
Matches	212; Conservative 0; Mismatches 264; Indels 0; Gaps 0;
Qy	19 GAAGGATPAACCTCTGTTCTTTCGGGAGAGCTGCAAAAGGCATAGCCACATAGTGT 78
Db	130 GCGAGGTCGTTTGGCACTCGGTGCGCGCATTAAGGATTGGCCATGATGATTT 189
Qy	79 TTGAAGCTATPAAGAGCTGGTATTAAGGGAGAGGCTTTTAAGGGGGGTAGCGCCGG 138
Db	190 ATTAGGTTTGAAGAAAGAAACGGATATCCCTGGAAGGGTGTATCCGACATCGCAGGT 249
Qy	139 GCAATCGTTTCGTCCTTTTATGCTCCAGACTACTCCCTGAAAGGATGTTACGCTTCTG 198
Db	250 TCGATAGTCGGAGCGCTTTTTCGATCGGGTATGTGCGCCGACCGCTCGAATTGGAAGCC 309
Qy	199 AAGAGGGTAAACTGGCTGAGCTGTTTAAGTTCAAGCCACTCTGAAAGGATGATAGGG 258
Db	310 GAAATTTTAGGTAAAAACCGATTTGGTCATTTAACCTTGTCACCAAGTGTATTATCAA 369
Qy	259 TGGGGAAGAGGCTATTAAGATTCCTTGAGAAATTCCTCTTACAGAGATAGAAAACTT 318
Db	370 GCGGAAAGCTGCAAAATTAATCAACGAAAGTGGCGGACGGCGATTCAGACGTTT 429
Qy	319 GAGATACCGAGATATATATATGCGCGACGAGATTATATCTGGGAGAGGGCTCTATACCTTCG 378
Db	430 CCCATCAAAATTTGCGCGCGCTGTCTATATTTTGAACCGGACAGAGCCGTGCTTTCAAT 489
Qy	379 GAAGGAGATTATATCCCGCACTTCTCGGACAGCTGTGCAATTCGCCGATATTTGAACCC 438
Db	490 CAAGGAAATGCGGGCAGAGCTGTGGCGCTTCCGCGCATTTCCAAATGTTCGAAACC 549
Qy	439 GTTAGATTAAGATTACTTGTCTGTTGACGAGGATATGTTAACACCTTCCGT 494
Db	550 GTTATCATCGCAGGACATATATGTTACGCGCGGTCTGTGCGAGCCGTCGCCGT 605
RESULT 8	
NMA122491/c	340806 bp DNA linear BCT 04-DEC-2000
LOCUS	NMA122491
DEFINITION	Neisseria meningitidis serogroup A strain Z2491 complete genome;
ACCESSION	segment 1/7.
VERSION	AL162752 AL157959
KEYWORDS	AL162752.2 GI:7378778
SOURCE	Neisseria meningitidis Z2491.
ORGANISM	Neisseria meningitidis Z2491
REFERENCE	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
AUTHORS	Neisseria.
TITLE	1 (bases 1 to 340806)
	Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C.,
	Klee, S. R., Morelli, G., Baeham, D., Brown, D., Chillingworth, T.,
	Davies, R. M., Davis, P., Devlin, K., Felwell, T., Hamlin, N.,
	Holtroff, S., Jagsels, K., Leather, S., Moul, S., Mungall, K.,
	Quail, M. A., Rajandream, M. A., Rutherford, K. M., Simmonds, S.,
	Skellon, J., Whitehead, S., Spratt, B. G. and Barrall, B. G.
	Complete DNA sequence of a serogroup A strain of Neisseria
	meningitidis Z2491

JOURNAL	Nature 404 (6777), 502-506 (2000)
MEDLINE	20222556
PUBMED	10761919
REFERENCE	2 (bases 1 to 340806)
AUTHORS	Parkhill,J.
TITLE	Direct Submission
JOURNAL	Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT	Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES	Location/Qualifiers
SOURCE	1..340806 /organism="Neisseria meningitidis Z2491" /strain="Z2491" /db_xref="taxon:122587" /note="serogroup: A" complement(38..358) /gene="NMA0001" complement(38..358) /gene="NMA0001" /note="NMA0001, unknown, len: 106 aa" /codon_start=1 /transl_table=1 /product="hypothetical protein NMA0001" /protein_id="CAB83321.1" /db_xref="GI:7378779" /db_xref="SPTRMBL:O9JX93" /translation="MSNTQTIRSDFTTSRQWALDIYANGTGTVALHWTDDPYKHVNDDEEFAVWDGVDMDHYRENGEBHTVRKSGDIFAGICTEHVAHPRGETRLIVIEKEEGSV"
gene	complement(414..2438) /gene="nuoL" complement(414..2438) /gene="nuoL"
CDS	/cfe_number="1.6.5.3" /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len: 674 aa; similar to many e.g. NUOL, RHOCA NADH dehydrogenase I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to entry PF00361 oxidored_q1, NADH-ubiquinone/plastoquinone (complex I), various chains, and to entry PF00662 oxidored q1' N, NADH-ubiquinone oxidoreductase (complex I), chain 5 N-terminus" /codon_start=1 /transl_table=1 /product="NADH dehydrogenase I chain L" /protein_id="CAB83322.1" /db_xref="GI:7378780" /db_xref="SWISS-PROT:O9JX92"
misc_feature	1035..1044 /note="Core DNA uptake sequence: gccctcgtaa" /label=DUS
misc_feature	complement(1116..2030) /gene="nuoL"
misc_feature	/note="Pfam match to entry PF00361 oxidored_q1, NADH-ubiquinone/plastoquinone (complex I), various chains, score 351.80, E-value 7.3e-102" complement(2061..2243) /gene="nuoL"

```

/note="pfam match to entry PF00662 oxidored_q1_N,
NADH-ubiquinone oxidoreductase (complex I), chain 5
N-terminus, score 77.20, E-value 3.5e-19"
/gene="NMA0003"
/complement(2487. .2828)
/EC_number="1.6.5.3"
/complement(2487. .2828)
/gene="NMA0003"
/note="NMA0003, unknown, len: 113 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0003"
/protein_id="CAB83323.1"
/db_xref="GI:7378781"
/db_xref="SPTREMBL:Q9JX91"
/translation="MMDKNQLQEFHFKAMLIYQALNLPQPKATKFLQIVNEFGGK
BAADKLLSTGKKTQTGFTLILSGGVHAKYSMEYLVLPKQWCDLFTERQAVARK
RLERVGVSPK"
/complement(2833. .2837)
/complement(2952. .3527)
/gene="NMA0004"
/complement(2952. .3527)
/gene="NMA0004"
/note="NMA0004, unknown, len: 191 aa; similar to
hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
fasta scores; E(): 0; 80.5% identity in 190 aa overlap,
and to proteins involved in cell cycle e.g. TR:054679
(EMBL:AF036487) Lactococcus lactis plasmid pNZ4000
putative mobilization protein (200 aa), fasta scores; E():
0, 55.1% identity in 167 aa overlap, and FIC ECOLI cell
filamentation protein FIC (200 aa), fasta scores; E():
0.093, 29.4% identity in 126 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0004"
/protein_id="CAB83324.1"
/db_xref="GI:7378782"
/db_xref="SPTREMBL:Q9JQR9"
/translation="MPSENPIGKTKMSIDEOSLNARLFESGDIIDRIEVTAGLQQ
IHRVFLGFLYDFAQIEDNISKGGFRPANAMYLKEALVKEQMPETFEIATKVE
IWIAPFLGNGRSTRITWLDVLKLNKLVVNMQVSKTLYLOAMERSPVNDLELRL
LKDNLTDDVNRREIIFKIGIOSYYEYEGK"
/complement(3516. .3525)
/gene="NMA0004"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(3561. .3866)
/gene="nuoK"
/complement(3561. .3866)
/gene="nuoK"
/EC_number="1.6.5.3"
/note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
101 aa; similar to many e.g. NUOK_RHOCA NADH dehydrogenase
I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
match to entry PF00420 oxidored_q2,
NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain K"
/protein_id="CAB83325.1"
/db_xref="GI:7378783"
/db_xref="SPTREMBL:Q9JQU9"
/translation="MITLHYLVIGALLFGISANGIFPNRNKLVLLMSTELMLAVN
FNFTAFSGHLDGTAGQIFVFFLVTAASEAIGLAIMVLVYRNQRTINVADLXLG"
/complement(3564. .3863)
/gene="nuoK"
/note="nuoK"
/note="Pfam match to entry PF00420 oxidored_q2,
NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
score 141.50, E-value 1.5e-38"
/complement(3863. .4534)
/gene="nuoJ"
/complement(3863. .4534)
/gene="nuoJ"

```

```

/EC_number="1.6.5.3"
/note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fasta scores; E():
9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFQILFIYFAVILYGAITVTAQNPVHAALHLVLPFCVSAM
LWMLMQAFLGTVTLVVYVQAVMFLFVWMLNIDIEEMRAGFWRHAPVAGVGTLL
AVALLILVNPKTDLAAEGLMKDIPADYNNIRDLGSRITYDYLPELAAVLLGLMV
AAIALVHRKTVNPKEMDPADQVKVRADQGRMLVKNQEAVKPQTESAESEVSDDLKPK
EEGKA"
/complement(3874. .3879)
/gene="nuoJ"
/complement(4034. .4525)
/gene="nuoJ"
/note="Pfam match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
score 77.90, E-value 2e-19"
/complement(4570. .5277)
/gene="NMA0007"
/complement(4570. .5277)
/gene="NMA0007"
/note="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"
/db_xref="GI:7378785"

```

```

Query Match      7.1%; Score 53.6; DB 1; Length 340806;
Best Local Similarity 44.5%; Pred. No. 0.00012;
Matches 212; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 19 GAAGAGATAAACCTCGTCTTTTCGGAGGAGCTGCAAGGCGATAGCCACATAGTGT 78
Db 215410 GCAGTGGTCGGTTTGGCACTCGGTGGCGCGCATCTAAAGGATTTGCCCATGTAGTATT 215351

QY 79 TTGAAAGCTATAAAACGAGCTCGGTATTAAGGGTGAGGGCTTTTAAGCGGGTGAGCCCGG 138
Db 215350 ATTAGGTTTTCGAAAGAAACGGTATTCTGTGAAGGTGGTTACCGGCACATCGCGAGT 215291

QY 139 GCAATCGTTTCGGTCTTTTANGCCTCAGGCTACTCCCTGAAAGGATGTTACGCTTCTG 198
Db 215290 TCGATAGTCGGCAGCGCTTTTTCATCGGTATGTGCGCGACCGCTCGAATTCGAAGCC 215231

QY 199 AAGAGGTTAACTGGCTGAAGCTGTTTAAGTTCAGCCACACCTCTGAAGGATGATAGGG 258
Db 215230 GAAATTTAGGTAAACCGATTTGGTTCGATTTCCTTCCACAGTGGTTTTATCAAA 215171

QY 259 TGGGAGAGGCTATAAGATTTCCTTCAGGAAGTTCTCCCTTACAGAGATAGAAAACTT 318
Db 215170 GCGGAAAGCTGCAAAATTACATCAACCGAAAGTGGCGGCGAGCGGATTTCAGCAGTTT 215111

QY 319 GAGATACCGAGCTATATATGCGCGACGGATTATTAATCTCGGGAAGGGCTCTATACCTTCG 378
Db 215110 CCCATCAAAATTTGCCCGCGCTTGTCTACTGATTTTGAACCGCGACGCGCTCGCTTTCAAT 215051

QY 379 GAAGGAGTTTAAATCCCGCACTTCTCGGCAGCTGTGCAATTCGCCGGATATTTCAACCC 438
Db 215050 CAAGGAATGCCGGGACGCTGTGCGGCTTCGCCGCCAATTTCCCAATGTGTTCACACCC 214991

QY 439 GTTAGTATAAGAAATTAATCTGCTCGTTGACGAGGATAGATTAAACAACCTTCCCTT 494
Db 214990 GTTATCATCGGCAGCATACATATGTTGACGGCGGTCTGTGCGACGCCGTGCCCT 214935

```

```

RESULT 9
LOCUS      A96526                      903 bp    DNA          linear    PAT 07-SBP-2000
DEFINITION Sequence 559 from Patent WO9924578.
ACCESSION  A96526
VERSION     A96526.1 GI:6780155
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 903)
AUTHORS     Pizza,M., Scariato,V., Rappuoli,R., Grandi,G. and Masignani,V.
TITLES      Neisserial antigens
JOURNAL     Patent: WO 9924578-A 559 20-MAY-1999;
            PIZZA MARIAGRAZIA (IT); SCARILATO VINCENZO (IT); RAPPUOLI RINO (IT);
            CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
FEATURES
  source     1..903
             /organism="unidentified"
             /db_xref="taxon:32644"
BASE COUNT  204 a 230 c 255 g 214 t
ORIGIN
Query Match      6.9%; Score 52; DB 6; Length 903;
Best Local Similarity 44.3%; Pired. No. 0.00028;
Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGTAACTCGTCTTTTGGAGAGAGCTGCAAGGCAATGACCATAGTGT 78
DB 130 GCGGTGTCGTTTGGACACTCGGTGGCGGCAATCAAGATTTGCCATGATATT 189
QY 79 TTGAAGCTATTAACAGAGCTCGGTATATAGGGGTATTAAGCGGGGTGAGCGCCG 138
DB 190 ATTAAAGTTTGAAGAAAGGATATTCCTGTGAAGGTGTTACCGGCACATCGCAGGT 249
QY 139 GCAATCGTTTGGTCTTTTATGCTTCAAGGCTACTCCCTGAAAGGATGTTCAAGCTTCG 198
DB 250 TCGATTTGGCGAGCCCTTTTGGATCGGGATATGCGCGACCGCTCGAATTTGAAAGCC 309
QY 199 AAGAGGTAACTGCTGAGAGCTGTTAAGTTACACCACTCTGAAAGGATTTGATAGG 258
DB 310 GAAATTTTGGGCAAAACCGATTTGTCGATTTAACTTTCACCAAGTGGTTTATCAAA 369
QY 259 TGGAGAAAGGCTATTAAGATTCTTGAAGAAATTCCTTACAGAAATAGAAACTT 318
DB 370 GGGGAAAGCTGCAAAATTTACATCAACGAAAGTGGCGGCGAGCATTCAGCATTT 429
QY 319 GAGATACCGAGATATATGCGGAGGATTTTACTCGGAAAGGCTCTTACTCTCG 378
DB 430 CCCATCAAAATTTGCGCGCGTGTCTACGATTTTGAACCGGAGGCGCGTCTTCAT 489
QY 379 GAAGGAGTTTATCCCGGCACTTCTCGGAGCTGTGCAATTCGCGCATTTGAACCC 438
DB 490 CAGGGAATTCGCGGAGGCTGTGCGCGCTTCCCGCCCATTTCCCAATGTGTTCAACCC 549
QY 439 GTTGAATTAAGATTACTGCTGTTGACGAGGATATGTTAACAACCTCCCGT 494
DB 550 GTTATCATCGGCGAGGATACATATGTTGACGCGCGGTCTGTGCGACGCCGTCGCGT 605

```

```

RESULT 10
LOCUS      AE002562/2                  9831 bp    DNA          linear    BCT 25-MAY-2000
DEFINITION Neisseria meningitidis serogroup B strain MC58 section 204 of 206
            of the complete genome.
ACCESSION  AE002562 AE002098
VERSION     AE002562.1 GI:7227392
KEYWORDS
SOURCE      Neisseria meningitidis MC58.
ORGANISM    Neisseria meningitidis MC58.
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1 (bases 1 to 9831)

```

```

AUTHORS     Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.P.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., Deboy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Clecko,A., Parksey,D.S., Blair,E.,
            Clifton,H., Clark,E.B., Cotton,M.D., Uterback,T.R., Khouri,H.,
            Qin,H., Vamathavan,J., Gill,J., Scariato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
TITLE       Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
JOURNAL     Science 287 (5459), 1809-1815 (2000)
MEDLINE     20175755
PUBMED      10710307
REFERENCE   2 (bases 1 to 9831)
AUTHORS     Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.P.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., Deboy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Clecko,A., Parksey,D.S., Blair,E.,
            Clifton,H., Clark,E.B., Cotton,M.D., Uterback,T.R., Khouri,H.,
            Qin,H., Vamathavan,J., Gill,J., Scariato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
TITLE       Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES
  source     1..9831
             /organism="Neisseria meningitidis MC58"
             /strain="MC58"
             /db_xref="taxon:122586"
             /note="serogroup: B"
             /gene="NMB2136"
             /gene="NMB2136"
             /gene="NMB2136"
             /note="similar to GB:AL009126 percent identity: 59.41;
             identified by sequence similarity; putative"
             /codon_start=1
             /product="peptide transporter"
             /protein_id="AA042444.1"
             /db_xref="GI:7227393"
             /translation="MSRRPAPPTGKTRFGHPFQSTLFIHLEMBRSFYMGQILITY
             LYTADKGIDKTLNAGIVAGVSGSVYSTLGAFAFVRWGAERTLPLSGIVNL
             GHVLAAPGIVGLIGLIFIALGSGGVKSTASMGVALYEDQEMRLPDRAGSIFPI
             AINTGFTGLGLQENIGFYHFGAAVAGMAFGIMRSLGRKMLPHPTVPHLSK
             GCGKTAAVGIALIALATATKGLVLDNFGSLISTVLAIVAPRLTLTPRVS
             DKRHTIATVPLFTICMPVAWVOITVATVTFDEVNNTGSPFVPAVMKMSGL
             WTLFSGIAAMNTKMRKOPKPIPKRAAAVPTGASFUDFVPPISSGTMPVAVRFL
             IVLAITIGELMISTIALISTKIAPIPKTQWALNPLARSLGFTLGVLPERGYONG
             DEIGFRLFLFYIGATGFLLLLVPKLNKLMBEGD"
             /gene="NMB2137"
             /gene="NMB2137"
             /note="hypothetical protein; identified by Glimmer2;
             putative"
             /codon_start=1
             /trnasl_table=11
             /product="hypothetical protein"
             /protein_id="AA042445.1"
             /db_xref="GI:7227394"
             /translation="MLKQPTRTTGGGRCVLFGLAPNGVWPAAYCYQMRGAPLPHLFT
             LTCAKAIAIGFAFCSTFRVTAAGR"
             /complement(3122..4425)
             /gene="NMB2138"
             /complement(3122..4425)
             /gene="NMB2138"
             /note="similar to GB:M38590 SB:P28353 PID:154276 percent
             identity: 80.11; identified by sequence similarity;
             putative"

```



Db	4741	CAGGGGATGCGGCGAGGCTGTGCGGCTTCCGCCGCAATCCCAATGTTTCCAAACC	4682
Qy	439	GTGAGTATTAAGATTACTTGTGCGTGAAGAGTATTAACAACCTTCCCGT	494
Db	4681	GTATATCATCGGACGACATATATGTTGACGGCGGTCTGTGCGAGCCCGTCCCGT	4626
RESULT 11			
LOCUS	AX044035/c	172325 bp	DNA linear
DEFINITION	Sequence 114 from Patent WO0066791.		PAT 24-NOV-2000
ACCESSION	AX044035		
VERSION	AX044035.1	GI:11342919	
KEYWORDS	Neisseria meningitidis.		
ORGANISM	Neisseria meningitidis.		
SOURCE	Bacteria: Proteobacteria; beta subdivision; Neisseriaceae;		
REFERENCE	1 (bases 1 to 172325)		
AUTHORS	Piazza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C., Masiagnani, V., Galeotti, C., Mora, M., Ratti, G., Scarselli, M., Scarlato, V., Rappunoli, R., Frazer, C.M. and Grandi, G.		
TITLE	Neisseria genomic sequences and methods of their use		
JOURNAL	Patent: WO 0066791-A 114 09-NOV-2000;		
FEATURES	CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)		
SOURCE	1. 172325		
	/organism="Neisseria meningitidis"		
	/db_xref="taxon:487"		
	/note="Sequence too long, cut in 8 pieces. -seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 2272325 172325 bases"		
BASE COUNT	43072 a 47583 c 41465 g 40205 t		
ORIGIN			
Query Match	6.9%; Score 52; DB 6; Length 172325;		
Best Local Similarity	44.3%; Pred. No. 0.0034;		
Matches 211; Conservative	0; Mismatches 265; Indels 0; Gaps 0;		
Qy	19	GAGAGATTAACCTCGTTCTTTCGGGAGAGCTGCAAGGGCATACCCCATAGGTGT	78
Db	153800	GCAATGTGCGTTTGGCACTCGGTGGCGCGCATTAAGAGATTGCCATGTAGGTAT	153741
Qy	79	TTGAAGCTTAACGAGCTCGGTATAGGGTGAAGGGCTTTAAGCGGGTGAGCCCGG	138
Db	153740	ATTAAAGTTTGAAGAAAGGATATCTGTGAAGTGGTTACCGGCACATCGCAGGT	153681
Qy	139	GCATTCGTTTCGTTTATGCTCAGGCTACTCCCTGAAAGGATGTTCAACCTCTG	198
Db	153680	TGCAATTCGCGACACCTTTTGTGATCGGTATGTCCCGGACCGCTCCATTTGAGACC	153621
Qy	199	AAGAGGTTAACTGCTGAAGCTGTTTAAGTTCAAGCAACCTGTGAAGGATTTAGGG	258
Db	153620	GAAATTTTGAAGCAAAACCGATTGTGCGATTAACTGTTCACACCGTGTATTATCAA	153561
Qy	259	TGGAGAGAGCTATTAAGATTCCTTGAAGAGTTCCTCCCTTACAGAGATTAAGAACT	318
Db	153560	GGCGAAAGCTGCAAAATTCATCAACCGAAAGTGGGGGAGAGATTCAGCAGTTT	153501
Qy	319	GAGATACCGACGATATATATGCGGAGCGAGTTTACTCGGAAAGGCTCTATACCTTCG	378
Db	153500	CCCATCAAAATTTGCGCGCGGTGCTACTGATTTGAAACCGGCAAGGCGCTTCCTTCAAT	153441
Qy	379	GAGAGAGATTAATCCCGGACCTTCGCGAGCTGTGCAATTCGCGCATATTGAAACC	438
Db	153440	CAGGGGAATGCGGCGAGGCTGTGCGGCTTCCGCCGCAATTCCTCAATGTTTCAACCC	153381
Qy	439	GTGAGTATTAAGATTACTTGTGCGTGAAGAGTATTAACAACCTTCCCGT	494

**AUTHORS** Nakasone, K., Masui, N., Takaki, Y., Sasaki, R., Maeno, G., Sasaki, R., Sakiyama, T., Hirama, C., Fuji, P. and Takami, H.  
**TITLE** Characterization and comparative study of the rrm operons of alkaliphilic *Bacillus halodurans* C-125  
**JOURNAL** Extremophiles 4 (4), 209-214 (2000)  
**MEDLINE** 20426005  
**PUBMED** 10972189  
**REFERENCE** 10 (sites)  
**AUTHORS** Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, P., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and Horikoshi, K.  
**TITLE** Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*  
**JOURNAL** Nucleic Acids Res. 28 (21), 4317-4331 (2000)  
**MEDLINE** 20512582  
**PUBMED** 11058132  
**REFERENCE** 11 (bases 1 to 300950)  
**AUTHORS** Takami, H. and Takaki, Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan  
**MEDLINE** (E-mail: takami@jamstec.go.jp)  
**JOURNAL** URL: <http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html>,  
 Tel: 81-468-67-3895, Fax: 81-468-66-6364

**FEATURES**  
 source

gene  
 CDS  
 /organism="Bacillus halodurans"  
 /db\_xref="taxon:86665"  
 /note="alkaliphile"  
 complement (179..2113)  
 /gene="spoVD"  
 complement (179..2113)  
 /gene="spoVD"  
 /note="BH2572"  
 /codon\_start=1  
 /transl\_table=11  
 /product="stage V sporulation protein (sporulation specific penicillin-binding protein) (spore cortex)"  
 /protein\_id="BAB06291.1"  
 /db\_xref="GI:10175193"  
 /translation="MRVSNVTVRRLIFVLLVGLATFLIIMRLGVVOPALGDMLTQ  
 AODSWSDVPEAKRGEITDRNDVLATNAPSILVPRVQKDAETAKLANVLQ  
 DRKAYELIKRESIVLNEPGRKISKDKANVRLKRLPGVIAEDSKHVPKGSYL  
 HVLGAGIDQGLTGLYDEQLKGEKHVSFPDAKQRLPNLADEYAPINGLNL  
 FLTIDSRVTIIRRELDIAEATYSPGATAINPNPTGILGMSRPHYNPNFREV  
 PEVYNQKPIVMQYEPGSTFKITLAALEENVEDLTDNRDHPGFIEVAGHRLCCK  
 KGGHSGTFLFLEWVENSNCNPFVILGERLKRDLFDYIEAFGQKGTGLDQGEAKGII  
 FNRDRIGLEQATFAGQSVSTPIQQAASVAAVNGVLYEPYAKDWVDPYTGCVY  
 ESSAPKMKRQIISBTSKEVRYALESVWARGTQGAFFVGVGKGTGTAOKAKOGRY  
 LNNHIVSFIFGAPADPOIIVVVAIDNPKDVTQFGVVAAPIVGKIIGDSIQALGIE  
 KRNQIEKERKNDPELVEPDLVGRTRKDLHESYYELKIDADGKDVVVAQAPPGT  
 KVAVGSTIRLYMDDTPTD"  
 complement (2295..4466)  
 /gene="pbbp"  
 complement (2295..4466)  
 /gene="pbbp"  
 /note="BH2573"  
 /codon\_start=1  
 /transl\_table=11  
 /product="penicillin-binding protein 2B (cell-division septum)"  
 /protein\_id="BAB06292.1"  
 /db\_xref="GI:10175194"  
 /translation="MDIKRAATNKRAMLVLFVVFVSLIARFLYLQIEKVGYSL  
 QAIABARWTSSEVLHKGRTIRDRNGDALAEVTSYVAVLDNSHOPAPERTQ  
 LLAHINNSPILLLELTTRDYOVELGAGANLTQQRDKIADLDLPGIYFTEERY  
 YPKHYASHLIGITDRDMGSRMGLESLDEYLKGDGSHIRFKKQGGIPLPSPESQI  
 EPKATGNDVYLTDNSIQALDQVMTQVEEYKPERIIAIVADPKKQGLIAMSNPFI  
 PNPEYQITNTMYAVSDRVEPGSTMKVFTLAAALEEVLNVEYQSGTYAIRNDTN  
 PPIRDHQKRGWGTISYLEAMQSSNVGFKSIALEKGLYELDFEGFGEIGTID  
 LPNEAASVYKAGALDAASTAFQGGTAVTPIQIQIQAATAIANDGKMKPVVVDRTVDS  
 ETGVEIEKEPVGVEFISKETAKEVRDILLETVTVTSSSGTGRPFYLEGDPVAGTGTA

gene  
 CDS

Query Match 6.9%; Score 52; DB 1; Length 300950;  
 Best Local Similarity 44.2%; Pred. No. 0.00035;  
 Matches 214; Conservative 0; Mismatches 270; Indels 0; Gaps 0;  
 Qy 32 TCGTTCTTTTCGGAGAGCTGCAAGGCGATAGCCACATAGTGTGTTTGAAGCTATAA 91  
 Db 16087 TGGCTTTAGGTCAGTGGAGCTCGTGGGTATGCACTATCGTGTGTTGAAGATTATAG 16028  
 Qy 92 ACGAGCTCGGTAAAGCGTGAAGCGCTTTAAAGCGGGGTAGCGCCGGGGAATCGTTTCGG 151  
 Db 16027 AACAGGAAAGATACCGCATAGACTATCTCCCGGTAGTAGTATGGAGCCCTTGTGGCCT 15968

OVNRPDPCYLNHGCKNLFSPFGMAPKDDPSVIVYVADRLSNLTNOVSGSEPVAKIEN  
 TIMNSLOVLINISPSVEHLKDEEGYELADPIGESARSAREELEQAMKVKYVLCEGD  
 TVGQOPYSGHKLLGERVILRTSESYTLPSMIGRWDLVKVAVLNVVNVNLPFGQG  
 FVVGQSIIEGDTITPDYVVELASPTFRKTKTKDEKDEDE"  
 complement (4529..4891)  
 /gene="BH2574"  
 complement (4529..4891)  
 /note="BH2574"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB06293.1"  
 /db\_xref="GI:10175195"  
 /translation="MNVARSIEHVRRQGPAPRRPRQRGKITKGEKMLGLM  
 TVAPLLFIGFSVNYATMYSVNREIHELEQTVAEQINDGLSLQVVELSSPERILKI  
 AKSLGKKJEDENVKVVQN"  
 complement (4939..5871)  
 /gene="BH2575"  
 complement (4939..5871)  
 /note="BH2575"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB06294.1"  
 /db\_xref="GI:10175196"  
 /translation="MPEHVTVLKNSVIGLNIPKDGIVVDTCLGAGHSQEIWKLTG  
 GGHLYAFDODQYALSHAKETLSPYRGSPTLIESNFRYMKLEELGVHVDGLVFLDG  
 VSPQLDEDERGFSYHEDAPLDMNRQSLSAYEVNQWDFIELMKLISRYGERFPA  
 KQAKIEQSRREKPIRTTIGELVDIIKEAPAPARTGTHGPAKRTFQAIRAVNDELG  
 AFEEALEAAILDTPAKRIVITFHSLEDRIKCKEMPRESKGPDPVPPGLVIPESYKA  
 TLKLTIKDPIVFTABEILNRRASAKLRIAEQ"  
 complement (5901..6332)  
 /gene="BH2576"  
 complement (5901..6332)  
 /note="BH2576"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB06295.1"  
 /db\_xref="GI:10175197"  
 /translation="MEMGYRNHNDKGMIIKPAKREBELGETFVVTGLDRCLFVYP  
 QVWKQESLKNLPKTKARAPTRFFSGCELCEDKQGRVNIASIPREPAQLKKE  
 CWWIGSVNRVEIWSKELWEEYFAESESSEFAENIVDFDL"  
 complement (6487..8103)  
 /gene="BH2577"  
 complement (6487..8103)  
 /note="BH2577"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB06296.1"  
 /db\_xref="GI:10175198"  
 /translation="MIVSELKLLPSKAAKDYLNNDMLSPFDYNIHOPTVFQOORLS  
 DLQOPYDRDALSALLSYQKRFAPHDKAAQVVKLKDPSRVVIVGGQAGLLTGPLY  
 TIYKAVTLLAREOERALGVVVFVFWFAGEDHLDREINAVPIEKNGRWSHRIEEK  
 RKXIAEAGLNKETLAKWLATVRSLPETEHPLIYERVKTLAGHSNTYTFDPFELL



[illegible]



JOURNAL MEDLINE 4 (sites)  
AUTHORS Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., and Shinagawa, H.  
TITLE Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12  
JOURNAL MEDLINE 21156231  
REFERENCE 5 (bases 1 to 291136)  
AUTHORS Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: kken@gen-info.osaka-u.ac.jp)  
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)  
COMMENT genome project.  
FEATURES  
source location/Qualifiers  
1. 291136  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/db\_xref="RIMD\_0509952"  
/db\_xref="taxon:83334"  
126. .380  
/gene="ECS1692"  
126. .380  
/gene="ECS1692"  
/note="similar to YMG\_ECOLI g11787445 percent identity 97 in 84 aa (Conserved in E. coli K-12)"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB35115.1"  
/db\_xref="GI:13361157"  
/translation="MTLWITIFGLIAGIATLIMPGRDGGFPLTCLIGVAVGSGWATWFGISISGFNLHSPFLVAVGALIVGVFLRRR"  
complement (430. .2400)  
/gene="ECS1693"  
complement (430. .2400)  
/gene="ECS1693"  
/note="probable TonB dependent outer membrane receptor, similar to TonB dependent outer membrane receptor PirA (Escherichia coli CFT073) g13661477[gblAAC61709.1] percent identity 97 in 656 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative TonB dependent outer membrane receptor"  
/protein\_id="BAB35116.1"  
/db\_xref="GI:13361158"  
/translation="MRLKHYLCTALSLAFTQAAVAORSDSLTWSSPVSSTTTVLDOPTMALDKONVAOLSVPGVVLQSGSRBQYKVGPDROVYPDPGVLYVYDGNLRLATLNNLGNVSKGYSILQGPOMOGALINITQKTPKLPASIRBQWRSQDNAYDMAHPASSSELYGVLSQGLDGLPHGVNNDLAKHGMMSSADKRGIVKGLPTRENDVTLTYITKDGKDNPPYSGNSGQSRPQWMPBPKESFYQGTTLANDRFTLSRLYRDTFENTLWYSLADLNKKSYSYSDYSQAGLQALADDMWDSYEAQKHEKDGSIITYDDNNOCAFVWQWKGKHFVNEBDTLASYDRKRPFTLKERYYTSKPAVNOALIVPOLKPERAGVDLTWNGAFTHDMGEVYVYVDRDALISHNIDADTIONSGTVDYSGLDAGIKGKICNIIIDVGLYLIHADAKRDIGITDLPLOTWATMLKPMPLSVTLSEPARSSVSNSGSOQAAGFAFTHIRAYTTLGHGSVANASVNNLEDTQYATSEGVSEGRNPAAGVETTP"  
complement (2426. .3280)  
/gene="ECS1694"  
complement (2426. .3280)  
/gene="ECS1694"  
/note="molybdenum transport protein, similar to molybdenum transport proteins e.g. g13661478[gblAAC61710.1] percent identity 91 in 284 aa"  
/evidence=not experimental  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="molybdenum transport protein"  
/protein\_id="BAB35117.1"  
/db\_xref="GI:13361159"  
/translation="MIFLSQAOIDALLLEDIQGDDLTTFALNIGHQGYIEFLRQCGCVSGVAVACRM/ITGLITIDVADSDGSQANAGRLIRAOGNAAALHQGWALQNVLEWSCGVSDYDQMLALIERYPDGNACTRKAIPGTRILASQALIAAGLIRACGAETIILFANHRHFLHNDQMSGAINQLRRHAEKIVVEADTPKBAIALAODVDLDFSPQATRIAOIALPSLAPHCITLALGGINLTLLKNYLDGCGRLRPFITSAPYAAADIKVSLQPAASIT"  
complement (3219. .3500)  
/gene="ECS1695"  
complement (3219. .3500)  
/gene="ECS1695"  
/note="unknown"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB35118.1"  
/db\_xref="GI:13361160"  
/translation="MPTLISVDKIANRITAPRNVLSRTSAGVLARLTWSVSGYINGINNKVLPSPLPAAGRSSGGLIAYRRHRCDDPFPSSGNNRCVTSGRYRGR"  
complement (3445. .4089)  
/gene="ECS1696"  
complement (3445. .4089)  
/gene="ECS1696"  
/note="unknown, similar to hypothetical proteins e.g. Orf2 (Escherichia coli CFT073) g13661479[gblAAC61711.1] percent identity 98 in 214 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB35119.1"  
/db\_xref="GI:13361161"  
/translation="MTLIDIPRADIYLOQLGLARTKTEPDRHQAERKAKENASPTBSYLOQLAKTIDLOQAQTLFDMGCGEYVSLADKLTITGVYSQGLNVAARRAALAKNDVHLIQRAWEDSDLEPCDIAVASSSTLVADRRQAMSLNQAIRLVYTHLVATSFVSPAIQRAAGREVIETLPNYIFALNVLYQWGIYAHVDIFRGQNCODNST"  
complement (4099. .4857)  
/gene="ECS1697"  
complement (4099. .4857)  
/gene="ECS1697"  
/note="probable ferric enterobactin transport, similar to ferric enterobactin transport ATP-binding protein (Escherichia coli CFT073) g13661480[gblAAC61712.1] percent identity 79 in 148 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative ferric enterobactin transport ATP-binding protein"  
/protein\_id="BAB35120.1"  
/db\_xref="GI:13361162"  
/translation="MTLALVRAHSAISGYRHPVLPDVSTLSOGTICCLGANGCKTLMSTLIGVPLIKGVLDLSPVOTLSHRQAAQALAPWADGIFARSVDWMLAPITGAFSPYKQERKATKBOLEKGIHILAAARPNTSSGGRQALALRALVQAPRLALIDEPASSLDSDHQIQLDITLQALKNNGMTWLSLTHPLAHNAIADSLIYVEPGRVTOQLPEQLTKKLAALYVNSADQIHNLIALISH"  
complement (4854. .5834)  
/gene="ECS1698"  
complement (4854. .5834)  
/gene="ECS1698"  
/note="probable ABC transporter protein (permease), similar to ABC transporter permeases e.g. (Haemophilus influenzae) g12501391[sp|O57130|YE71\_HAin percent identity 40 in 323 aa"  
/codon\_start=1  
transport proteins e.g. g13661478[gblAAC61710.1] percent identity 91 in 284 aa"  
/evidence=not experimental  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="molybdenum transport protein"  
/protein\_id="BAB35117.1"  
/db\_xref="GI:13361159"  
/translation="MIFLSQAOIDALLLEDIQGDDLTTFALNIGHQGYIEFLRQCGCVSGVAVACRM/ITGLITIDVADSDGSQANAGRLIRAOGNAAALHQGWALQNVLEWSCGVSDYDQMLALIERYPDGNACTRKAIPGTRILASQALIAAGLIRACGAETIILFANHRHFLHNDQMSGAINQLRRHAEKIVVEADTPKBAIALAODVDLDFSPQATRIAOIALPSLAPHCITLALGGINLTLLKNYLDGCGRLRPFITSAPYAAADIKVSLQPAASIT"  
complement (3219. .3500)  
/gene="ECS1695"  
complement (3219. .3500)  
/gene="ECS1695"  
/note="unknown"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB35118.1"  
/db\_xref="GI:13361160"  
/translation="MPTLISVDKIANRITAPRNVLSRTSAGVLARLTWSVSGYINGINNKVLPSPLPAAGRSSGGLIAYRRHRCDDPFPSSGNNRCVTSGRYRGR"  
complement (3445. .4089)  
/gene="ECS1696"  
complement (3445. .4089)  
/gene="ECS1696"  
/note="unknown, similar to hypothetical proteins e.g. Orf2 (Escherichia coli CFT073) g13661479[gblAAC61711.1] percent identity 98 in 214 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB35119.1"  
/db\_xref="GI:13361161"  
/translation="MTLIDIPRADIYLOQLGLARTKTEPDRHQAERKAKENASPTBSYLOQLAKTIDLOQAQTLFDMGCGEYVSLADKLTITGVYSQGLNVAARRAALAKNDVHLIQRAWEDSDLEPCDIAVASSSTLVADRRQAMSLNQAIRLVYTHLVATSFVSPAIQRAAGREVIETLPNYIFALNVLYQWGIYAHVDIFRGQNCODNST"  
complement (4099. .4857)  
/gene="ECS1697"  
complement (4099. .4857)  
/gene="ECS1697"  
/note="probable ferric enterobactin transport, similar to ferric enterobactin transport ATP-binding protein (Escherichia coli CFT073) g13661480[gblAAC61712.1] percent identity 79 in 148 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative ferric enterobactin transport ATP-binding protein"  
/protein\_id="BAB35120.1"  
/db\_xref="GI:13361162"  
/translation="MTLALVRAHSAISGYRHPVLPDVSTLSOGTICCLGANGCKTLMSTLIGVPLIKGVLDLSPVOTLSHRQAAQALAPWADGIFARSVDWMLAPITGAFSPYKQERKATKBOLEKGIHILAAARPNTSSGGRQALALRALVQAPRLALIDEPASSLDSDHQIQLDITLQALKNNGMTWLSLTHPLAHNAIADSLIYVEPGRVTOQLPEQLTKKLAALYVNSADQIHNLIALISH"  
complement (4854. .5834)  
/gene="ECS1698"  
complement (4854. .5834)  
/gene="ECS1698"  
/note="probable ABC transporter protein (permease), similar to ABC transporter permeases e.g. (Haemophilus influenzae) g12501391[sp|O57130|YE71\_HAin percent identity 40 in 323 aa"  
/codon\_start=1

```

/evidence-not experimental
/transl_table=11
/product="putative ABC transporter permease protein"
/protein_id="BAB35121.1"
/db_xref="GI:13361163"
/translation="NR1VNGCILLAAISITPAVSGAYHLDMOQLLALILROBNFV
QEQIVFQWRVPRILAAALLGAALAGAGTTCQMLRNPVSPDILGVSAAGLACTA
ILWGLSVITQYAFPCGGLMVAVGWLITRRVTRHDPILITLVGLALGTLCGAGISL
IKTLADPTQLPSITFWLGGSLSTVLRDLCAAPILIGLSLPLFLWRMLLITLSD
DEARSLGNVTRLPVLIVCATLITASTVAIAGLIGWGLVVPVHIALTLGHNHQQLL
PMANCTGAILLLTDTLARSIGTTEPLIGLITAFVCAFPFLLLLRGRQ"
gene      948 a 916 c 843 g 946 t
CDS
complement(5834..6856)
/genes="Ece1699"
/notes="probable ABC transporter, weakly similar to iron
(iii) ABC transporter, ATP-binding protein [Pyrococcus
Query Match      6.4%; Score 49.8; DB 1; Length 291136;
Best Local Similarity 52.7%; Pred. No. 0.0016;
Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 340 GCGACGGATTATATCTCGGGAAGGCTCTATACCTCTCGGAAGGAGTTTAATCCCGCA 399
Db 47183 GCCACCAATTTAAGTACTGGCGTGAATATGTTTACTGAAGGCGATCTCCATCTTGT 47242
QY 400 CTCTCGGAGCTGTGCANTTCCCGGCATATTGTAACCCGTTGATTAAGAAATPACTTG 459
Db 47243 ATTGCGCATCATGCGAGTATTCAGGACTAATGCGACCCGTTGCGCATACGGCTACTGG 47302
QY 460 CTGTTGACGAGGTATAGTTAAACCTTCGCTTGAGCCCTTTTCAGAAAGCGTATT 519
Db 47303 CTGTTGATGGGAGTCTGTAAACCAATCTTTCGCTCAGCGGTGATGGTGTCT 47362
QY 520 CCCACCGTTTGGCTGATGCTCTTC 544
Db 47363 GATATTGTGATAGCGGTGACCTGC 47387

```

```

RESULT 15
ECOTGSA/c
LOCUS      ECOTGSA      3653 bp      DNA      linear      BCT 18-MAR-1994
DEFINITION Escherichia coli tgs gene, partial cds.
ACCESSION M64675
VERSION M64675.1 GI:402693
KEYWORDS   tgs gene.
SOURCE     Escherichia coli (strain K-12) DNA.
ORGANISM   Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE 1 (bases 1 to 3653)
AUTHORS   Bosl,M. and Kersten,H.
TITLE      Organization and functions of genes in the upstream region of tyrT
            of Escherichia coli: phenotypes of mutants with partial deletion of
            a new gene (tgs)
JOURNAL    J. Bacteriol. 176 (1), 221-231 (1994)
MEDLINE    94110230
PUBMED     8282700
FEATURES   source
            Location/Qualifiers
            1..3653
            /organism="Escherichia coli"
            /strain="K-12"
            /db_xref="taxon:562"
            1154..1159
            TATA_signal
            -35_signal
            1176..1181
            TATA_signal
            gene
            2213..2218
            2524..3479
            /gene="tgs"
            -35_signal
            2524..2529
            /gene="tgs"
            TATA_signal
            2548..2553
            /gene="tgs"
            CDS
            <2687..3479

```

```

/gene="tgs"
/codon_start=2
/transl_table=11
/protein_id="AAA16860.1"
/db_xref="GI:402694"
/translation="GLIARIITNICYKHELNIVQNEFVDHRTGRFFMRTELEGI FND
TLLADLSALPEGSVRELNPAGRRRIIVLTKEAHCLGDLIMKANYGGLDVRVIAAVIG
NHDITRSUVERFDIPFELVSHGELTRNEHQKADADAQOPDVVVLAKYMRVLTPEF
VARPNKTIINIHSLFPAFICARPYHOAVERGVKICATARYNDNLDEGPIIMQDVI
HVDHTYTAEDMMRAGROVEKNVLSRALYKVLQAORVYVYGNRTLL"
BASE COUNT      948 a 916 c 843 g 946 t
ORIGIN
Query Match      6.4%; Score 48.2; DB 1; Length 3653;
Best Local Similarity 52.2%; Pred. No. 0.0041;
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 340 GCGACGGATTATATCTCGGGAAGGCTCTATACCTCTCGGAAGGAGTTTAATCCCGCA 399
Db 1680 GCCACCAATTTAAGTACCGGACGTGAATATGTTTACTGAAGCGGATCTCCATCTTGT 1621
QY 400 CTCTCGGCGAGCTGTGCAATTCGCGCATATTGAAACCCGTTGAGTATAAGAAATTAATTG 459
Db 1620 ATTGCGCATCATGCGAGTATTCAGGACTCATGCACTGTTGCACATAACGGCTACTGG 1561
QY 460 CTGCTTGACGAGGTATAGTTAAACCTTCGCTTGAGCCCTTTTCAGAAAGCGGTATT 519
Db 1560 CTGCTTGATGGAGCAGTCGTTAAACCAATCTCTATTTCCCTCACGCGTGCATTGGGGCT 1501
QY 520 CCCACCGTTTGGCTGATGCTCTTC 544
Db 1500 GATATTGTGATAGCGGTGACCTGC 1476

```

Search completed: June 19, 2003, 14:02:48  
Job time : 1454 secs